GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 29, 2004, 23:39:31 ; Search time 125 Seconds
(without alignments)
12886.115 Million cell updates/sec

Title: US-10-032-996-6
Perfect score: 5739
Sequence: 1 gcgccctgagctccgcctcc.....aaaaaaatctttgaagggac 3121

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10032996/runat_29012004_160249_24126/app_query.fasta_1.3271
-DB=SPTREMBL_23 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10032996_@CGN_1_1_168_@runat_29012004_160249_24126 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

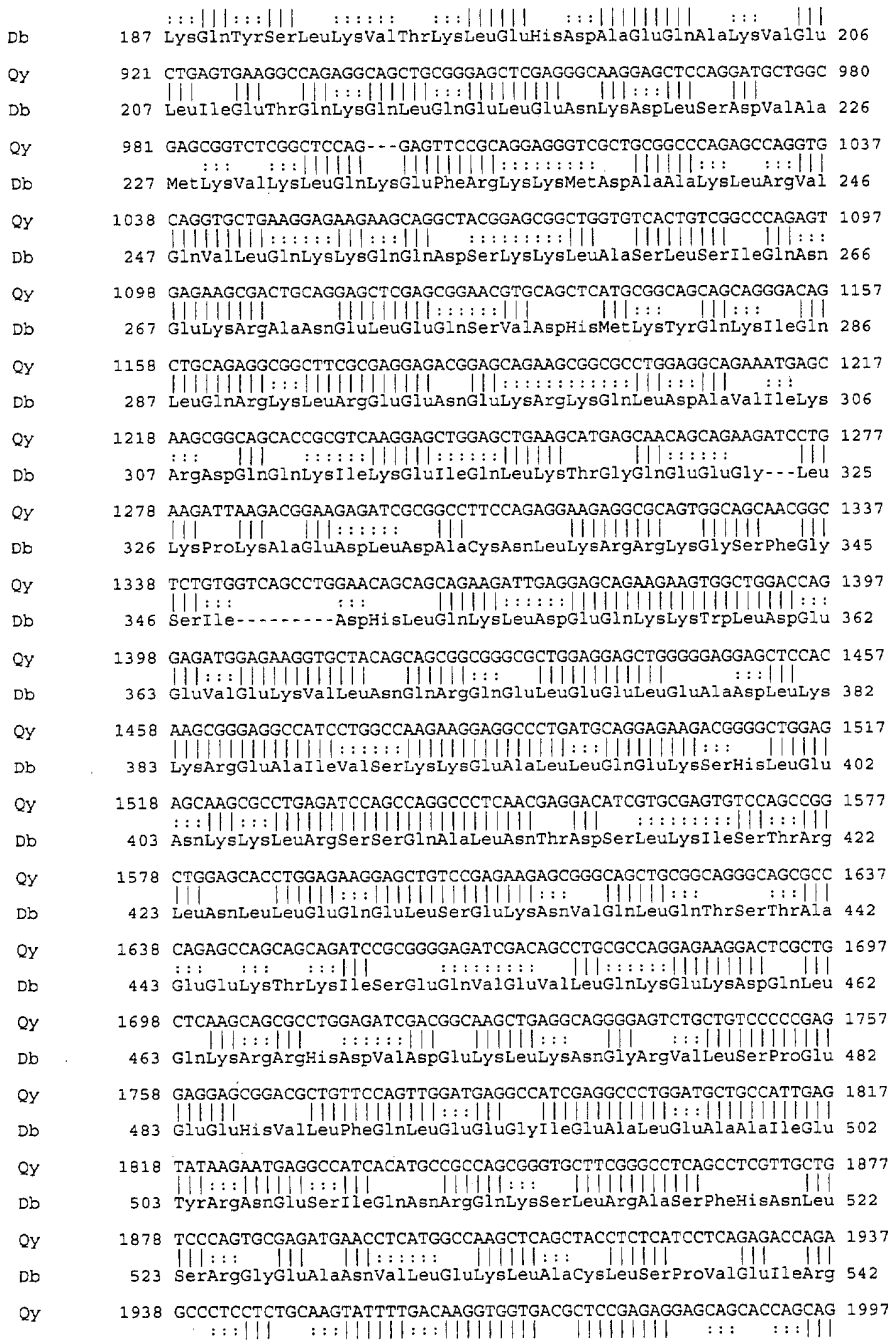
Result No.	Score	% Query Match	Length	DB	ID	Description
1	2334	40.7	468	4	Q8IW72	Q8iw72 homo sapien
2	1250.5	21.8	833	4	Q9UF54	Q9uf54 homo sapien
3	1243.5	21.7	669	11	Q8BMB9	Q8bmb9 mus musculu
c 4	626.5	11.0	676	6	Q95JC9	Q95jc9 sus scrofa
5	598	10.4	1953	5	Q9BIT7	Q9bit7 nephila mad
c 6	560.5	9.8	1450	13	Q9YIB4	Q9yib4 cynops pyrr
c 7	560.5	9.8	1453	11	Q63079	Q63079 rattus norv
c 8	553	9.7	950	5	Q9VE45	Q9ve45 drosophila
c 9	548.5	9.6	566	6	Q95JD1	Q95jd1 sus scrofa
c 10	541.5	9.5	1099	5	Q8IR84	Q8ir84 drosophila
c 11	540.5	9.5	1461	4	O76045	O76045 homo sapien
c 12	540.5	9.5	1464	4	Q8N473	Q8n473 homo sapien
c 13	533	9.3	1447	13	Q9IB91	Q9ib91 xenopus lae
c 14	532.5	9.3	1464	11	Q8BLW4	Q8blw4 mus musculu
c 15	531.5	9.3	1464	11	Q8BKY2	Q8bky2 mus musculu
16	527	9.2	1447	13	Q9IB91	Q9ib91 xenopus lae
c 17	523	9.2	1445	13	O93251	O93251 rana catesb
18	522	9.1	1445	13	O93251	O93251 rana catesb
19	518.5	9.0	1449	13	Q910C0	Q910c0 oncorhynchu
20	517	9.0	1487	4	Q14047	Q14047 homo sapien
21	511	8.9	1160	4	Q14046	Q14046 homo sapien
c 22	511	8.9	1420	13	Q90W37	Q90w37 gallus gall
23	509.5	8.9	1461	4	O76045	O76045 homo sapien
24	509.5	8.9	1464	4	Q8N473	Q8n473 homo sapien
25	509	8.9	774	12	O41971	O41971 murid herpe
c 26	509	8.9	2715	5	O61603	O61603 drosophila
c 27	508	8.9	1442	11	Q62031	Q62031 mus musculu
c 28	508	8.9	1442	11	Q62033	Q62033 mus musculu
c 29	508	8.9	1459	11	Q62032	Q62032 mus musculu
c 30	507	8.9	1487	4	Q14047	Q14047 homo sapien
c 31	506	8.9	1160	4	Q14046	Q14046 homo sapien
32	505.5	8.8	1346	13	Q8UUJ3	Q8uuj3 oncorhynchu
33	504.5	8.8	1352	13	Q8UUJ4	Q8uuj4 oncorhynchu
34	504.5	8.8	1453	11	Q63079	Q63079 rattus norv
c 35	504.5	8.8	2703	5	Q9VEG7	Q9veg7 drosophila
c 36	504.5	8.8	2716	5	Q8IN94	Q8in94 drosophila
37	504	8.8	1458	13	Q910B9	Q910b9 oncorhynchu
c 38	504	8.8	3084	12	Q8UZ11	Q8uz11 pseudorabie
c 39	502	8.8	1449	13	Q910C0	Q910c0 oncorhynchu
c 40	501.5	8.8	1953	5	Q9BIT7	Q9bit7 nephila mad
c 41	501	8.8	1487	6	O77753	O77753 canis famil
42	500.5	8.7	1418	6	Q28396	Q28396 equus cabal
43	500.5	8.7	1439	5	O97406	O97406 haliotis di
44	498.5	8.7	1414	5	Q26634	Q26634 strongyloce
45	498	8.7	1442	11	Q62031	Q62031 mus musculu

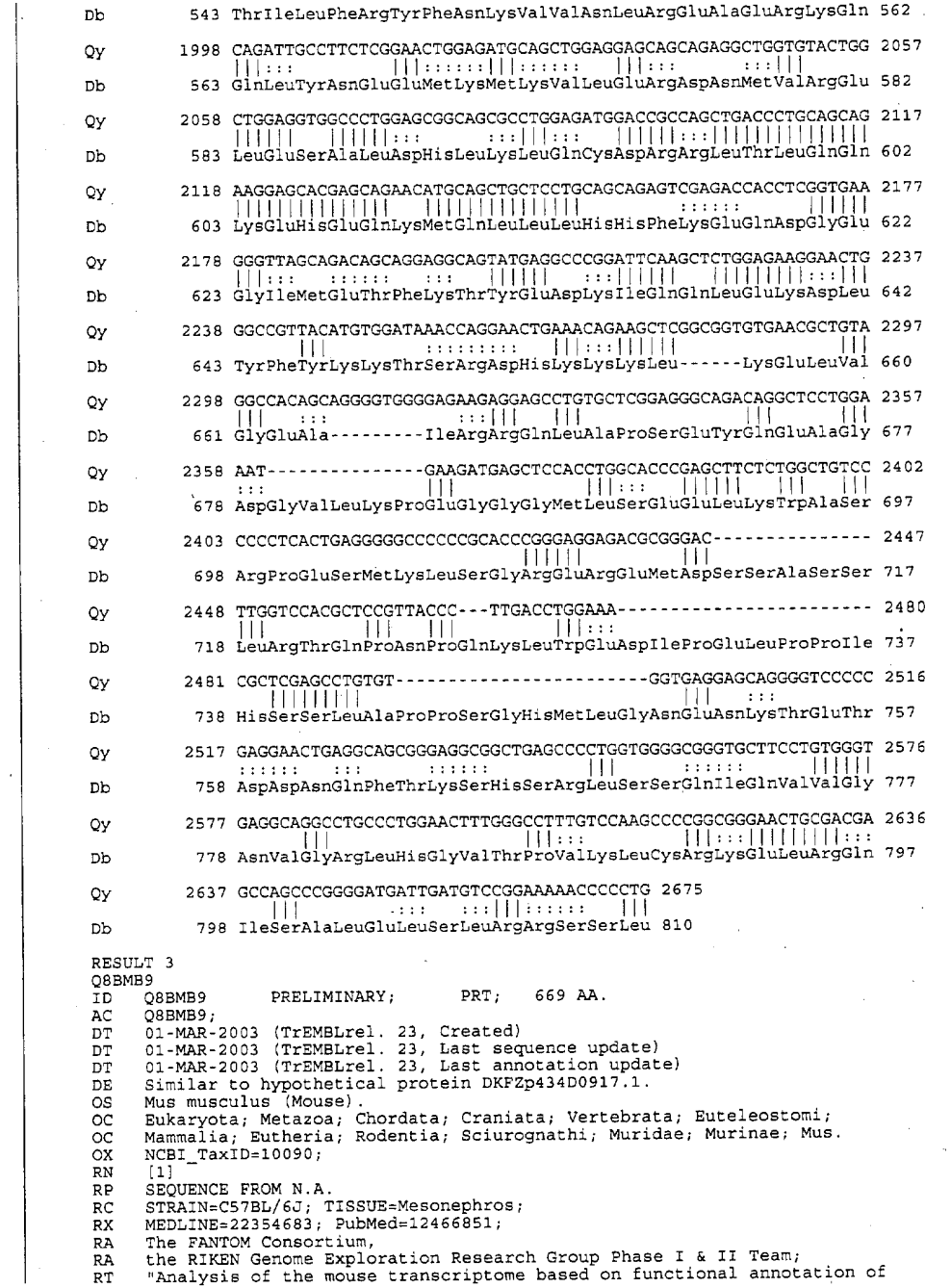
ALIGNMENTS

RESULT 1
Q8IW72
ID Q8IW72 PRELIMINARY; PRT; 468 AA.
AC Q8IW72;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strausberg R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC040878; AAH40878.1; -.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 468 AA; 54032 MW; CE4172079C033B92 CRC64;

Alignment Scores:
Pred. No.: 2.32e-100 Length: 468
Score: 2334.00 Matches: 466

```
Percent Similarity:      99.57%          Conservative:   0
Best Local Similarity:   99.57%          Mismatches:     2
Query Match:             40.67%          Indels:         0
DB:                      4               Gaps:           0

US-10-032-996-6 (1-3121) x Q8IW72 (1-468)

Qy       1272 ATCCTGAAGATTAAGACGGAAGAGATCGCGGCCTTCCAGAGGAAGAGGCGCAGTGGCAGC 1331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 IleLeuLysIleLysThrGluGluIleAlaAlaPheGlnArgLysArgArgSerGlySer 20

Qy       1332 AACGGCTCTGTGGTCAGCCTGGAACAGCAGCAGAAGATTGAGGAGCAGAAGAAGTGGCTG 1391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AsnGlySerValValSerLeuGluGlnGlnGlnLysIleGluGluGlnLysLysTrpLeu 40

Qy       1392 GACCAGGAGATGGAGAAGGTGCTACAGCAGCGGCGGGCGCTGGAGGAGCTGGGGGAGGAG 1451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AspGlnGluMetGluLysValLeuGlnGlnArgArgAlaLeuGluGluLeuGlyGluGlu 60

Qy       1452 CTCCACAAGCGGGAGGCCATCCTGGCCAAGAAGGAGGCCCTGATGCAGGAGAAGACGGGG 1511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuHisLysArgGluAlaIleLeuAlaLysLysGluAlaLeuMetGlnGluLysThrGly 80

Qy       1512 CTGGAGAGCAAGCGCCTGAGATCCAGCCAGGCCCTCAACGAGGACATCGTGCGAGTGTCC 1571
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuGluIleLysArgLeuArgSerSerGlnAlaLeuAsnGluAspIleValArgValSer 100

Qy       1572 AGCCGGCTGGAGCACCTGGAGAAGGAGCTGTCCGAGAAGAGCGGGCAGCTGCGGCAGGGC 1631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerArgLeuGluHisLeuGluLysGluLeuSerGluLysSerGlyGlnLeuArgGlnGly 120

Qy       1632 AGCGCCCAGAGCCAGCAGCAGATCCGCGGGGAGATCGACAGCCTGCGCCAGGAGAAGGAC 1691
              ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        121 SerAlaGlnSerGlnGlnGlnIleArgArgGluIleAspSerLeuArgGlnGluLysAsp 140

Qy       1692 TCGCTGCTCAAGCAGCGCCTGGAGATCGACGGCAAGCTGAGGCAGGGGGAGTCTGCTGTCC 1751
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerLeuLeuLysGlnArgLeuGluIleAspGlyLysLeuArgGlnGlySerLeuLeuSer 160

Qy       1752 CCCGAGGAGGAGCGGACGCTGTTCCAGTTGGATGAGGCCATCGAGGCCCTGGATGCTGCC 1811
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ProGluGluGluArgThrLeuPheGlnLeuAspGluAlaIleGluAlaLeuAspAlaAla 180

Qy       1812 ATTGAGTATAAGAATGAGGCCATCACATGCCGCCAGCGGGTGCTTCGGGCCTCAGCCTCG 1871
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluTyrLysAsnGluAlaIleThrCysArgGlnArgValLeuArgAlaSerAlaSer 200

Qy       1872 TTGCTGTCCCAGTGCGAGATGAACCTCATGGCCAAGCTCAGCTACCTCTCATCCTCAGAG 1931
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuLeuSerGlnCysGluMetAsnLeuMetAlaLysLeuSerTyrLeuSerSerSerGlu 220

Qy       1932 ACCAGAGCCCTCCTCTGCAAGTATTTTGACAAGGTGGTGACGCTCCGAGAGGAGCAGCAC 1991
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ThrArgAlaLeuLeuCysLysTyrPheAspLysValValThrLeuArgGluGluGlnHis 240

Qy       1992 CAGCAGCAGATTGCCTTCTCGGAACTGGAGATGCAGCTGGAGGAGCAGCAGAGGCTGGTG 2051
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GlnGlnGlnIleAlaPheSerGluLeuGluMetGlnLeuGluGluGlnGlnArgLeuVal 260

Qy       2052 TACTGGCTGGAGGTGGCCCTGGAGCGGCAGCGCCTGGAGATGGACCGCCAGCTGACCCTG 2111
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 TyrTrpLeuGluValAlaLeuGluArgGlnArgLeuGluMetAspArgGlnLeuThrLeu 280

Qy       2112 CAGCAGAAGGAGCACGAGCAGAACATGCAGCTGCTCCTGCAGCAGAGTCGAGACCACCTC 2171
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GlnGlnLysGluHisGluGlnAsnMetGlnLeuLeuLeuGlnGlnSerArgAspHisLeu 300

Qy       2172 GGTGAAGGGTTAGCAGACAGCAGGAGGCAGTATGAGGCCCGGATTCAAGCTCTGGAGAAG 2231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GlyGluGlyLeuAlaAspSerArgArgGlnTyrGluAlaArgIleGlnAlaLeuGluLys 320

Qy       2232 GAACTGGGCCGTTACATGTGGATAAACCAGGAACTGAAACAGAAGCTCGGCGGTGTGAAC 2291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        321 GluLeuGlyArgTyrMetTrpIleAsnGlnGluLeuLysGlnLysLeuGlyGlyValAsn 340

Qy       2292 GCTGTAGGCCACAGCAGGGGTGGGGAGAAGAGGAGCCTGTGCTCGGAGGGCAGACAGGCT 2351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AlaValGlyHisSerArgGlyGlyGluLysArgSerLeuCysSerGluGlyArgGlnAla 360

Qy       2352 CCTGGAAATGAAGATGAGCTCCACCTGGCACCCGAGCTTCTCTGGCTGTCCCCCCCTCACT 2411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ProGlyAsnGluAspGluLeuHisLeuAlaProGluLeuLeuTrpLeuSerProLeuThr 380

Qy       2412 GAGGGGGCCCCCCCGCACCCGGGAGGAGACGCGGGACTTGGTCCACGCTCCGTTACCCTTG 2471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GluGlyAlaProArgThrArgGluGluThrArgAspLeuValHisAlaProLeuProLeu 400

Qy       2472 ACCTGGAAACGCTCGAGCCTGTGTGGTGAGGAGCAGGGGTCCCCCGAGGAACTGAGGCAG 2531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ThrTrpLysArgSerSerLeuCysGlyGluGluGlnGlySerProGluGluLeuArgGln 420

Qy       2532 CGGGAGGCGGCTGAGCCCCTGGTGGGGCGGGTGCTTCCTGTGGGTGAGGCAGGCCTGCCC 2591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ArgGluAlaAlaGluProLeuValGlyArgValLeuProValGlyGluAlaGlyLeuPro 440

Qy       2592 TGGAACTTTGGGCCTTTGTCCAAGCCCCGGCGGGAACTGCGACGAGCCAGCCCGGGGATG 2651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 TrpAsnPheGlyProLeuSerLysProArgArgGluLeuArgArgAlaSerProGlyMet 460

Qy       2652 ATTGATGTCCGGAAAAACCCCCTG 2675
              ||||||||||||||||||||||||
Db        461 IleAspValArgLysAsnProLeu 468

RESULT 2
Q9UF54
ID   Q9UF54      PRELIMINARY;      PRT;   833 AA.
AC   Q9UF54;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP434D0917.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,
RA   Wiemann S.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133654; CAB63770.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   833 AA;  96688 MW;  72D59966CFB177F3 CRC64;

Alignment Scores:
Pred. No.:               2.79e-50        Length:          833
Score:                   1250.50         Matches:         291
Percent Similarity:      65.23%          Conservative:    148
Best Local Similarity:   43.24%          Mismatches:      197
Query Match:             21.79%          Indels:          37
DB:                      4               Gaps:            10

US-10-032-996-6 (1-3121) x Q9UF54 (1-833)

Qy        741 TCAGAGTGGCGGCTGGCCCAGGCCCAGCAGAAGATCCGGGAGCTGGCTATCAACATCCGC 800
              ||||||     |||  :::|||:::|||||||:::|||:::|||   :::      |||
Db        147 SerGluArgIleLeuThrGluAlaLysGlnLysMetArgAspLeuGlnLeuThrSerArg 166

Qy        801 ATGAAGGAGGAGCTTATTGGCGAGCTGGTCCGCACAGGAAAGGCAGCTCAGGCCCTGAAC 860
              ||||||||:::|||||   ||||||:::::|||||       |||:::::::::::::
Db        167 MetLysGluAspLeuIleLysGluLeuIleLysThrGlyAsnAspAlaLysSerValSer 186

Qy        861 CGCCAGCACAGCCAGCGTATCCGGGAGCTGGAGCAGGAGGCAGAGCAGGTGCGGGCCGAG 920
```






```
RESULT 3
Q8BMB9
ID   Q8BMB9      PRELIMINARY;      PRT;   669 AA.
AC   Q8BMB9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to hypothetical protein DKFZp434D0917.1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Mesonephros;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
```

RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK032912; BAC28083.1; -.
KW Hypothetical protein.
SQ SEQUENCE 669 AA; 77554 MW; 4D27F3C92D0101FE CRC64;

Alignment Scores:

Pred. No.:	5.82e-50	Length:	669
Score:	1243.50	Matches:	293
Percent Similarity:	63.40%	Conservative:	140
Best Local Similarity:	42.90%	Mismatches:	203
Query Match:	21.67%	Indels:	47
DB:	11	Gaps:	12

US-10-032-996-6 (1-3121) x Q8BMB9 (1-669)

```
Qy        774 ATCCGGGAGCTGGCTATCAACATCCGCATGAAGGAGGAGCTTATTGGCGAGCTGGTCCGC 833
Db          1 MetArgGluLeuThrIleAsnIleArgMetLysGluAspLeuIleLysGluLeuIleLys 20

Qy        834 ACAGGAAAGGCAGCTCAGGCCCTGAACCGCCAGCACAGCCAGCGTATCCGGGAGCTGGAG 893
Db         21 ThrGlyAsnAsnAlaLysSerValSerArgGlnTyrSerLeuLysValThrLysLeuGlu 40

Qy        894 CAGGAGGCAGAGCAGGTGCGGGCCGAGCTGAGTGAAGGCCAGAGGCAGCTGCGGGAGCTC 953
Db         41 HisGluAlaGluGlnAlaLysValGluLeuThrGluThrArgLysGlnLeuGlnGluLeu 60

Qy        954 GAGGGCAAGGAGCTCCAGGATGCTGGCGAGCGGTCTCGGCTCCAG---GAGTTCCGCAGG 1010
Db         61 GluSerLysAspLeuSerAspValAlaLeuLysValLysLeuGlnLysGluPheArgLys 80

Qy       1011 AGGGTCGCTGCGGCCCAGAGCCAGGTGCAGGTGCTGAAGGAGAAGAAGCAGGCTACGGAG 1070
Db         81 LysMetAspAlaAlaLysMetArgValGlnValLeuGlnLysLysGlnGlnAspSerLys 100

Qy       1071 CGGCTGGTGTCACTGTCGGCCCAGAGTGAGAAGCGACTGCAGGAGCTCGAGCGGAACGTG 1130
Db        101 LysLeuAlaSerLeuSerIleGlnAsnGluLysArgAlaSerGluLeuGluHisAsnVal 120

Qy       1131 CAGCTCATGCGGCAGCAGCAGGGACAGCTGCAGAGGCGGCTTCGCGAGGAGACGGAGCAG 1190
Db        121 AspHisLeuLysTyrGlnLysValGlnLeuGlnArgArgLeuArgGluGluGlyGluLys 140

Qy       1191 AAGCGGCGCCTGGAGGCAGAAATGAGCAAGCGGCAGCACCGCGTCAAGGAGCTGGAGCTG 1250
Db        141 LysLysGlnLeuAspAlaGluIleLysArgAspGlnGlnLysIleLysGluLeuGlnLeu 160

Qy       1251 AAGCATGAGCAACAGCAGAAGATCCTGAAGATTAAGACGGAAGAGATCGCGGCCTTCCAG 1310
Db        161 LysAlaGlyGlnGlyGluGly---LeuAsnProLysAlaGluAspGlnAspGlyPheAsn 179

Qy       1311 AGGAAGAGGCGCAGTGGCAGCAACGGCTCTGTGGTCAGCCTGGAACAGCAGCAGAAGATT 1370
Db        180 LeuAsnArgArg---------LysSerProPheArgSerGlyAspGlnPheGlnLysLeu 196

Qy       1371 GAGGAGCAGAAGAAGTGGCTGGACCAGGAGATGGAGAAGGTGCTACAGCAGCGGCGGGCG 1430
Db        197 AspGluGlnArgLysTrpLeuAspGluGluValGluLysValLeuSerGlnArgGlnGlu 216

Qy       1431 CTGGAGGAGCTGGGGGAGGAGCTCCACAAGCGGGAGGCCATCCTGGCCAAGAAGGAGGCC 1490
Db        217 LeuGluMetLeuGluGluAspLeuLysLysArgGluAlaIleValSerLysLysGluAla 236

Qy       1491 CTGATGCAGGAGAAGACGGGGCTGGAGAGCAAGCGCCTGAGATCCAGCCAGGCCCTCAAC 1550
Db        237 LeuLeuGlnGluLysSerLeuLeuGluAsnLysLysLeuArgSerSerGlnAlaLeuSer 256

Qy       1551 GAGGACATCGTGCGAGTGTCCAGCCGGCTGGAGCACCTGGAGAAGGAGCTGTCCGAGAAG 1610
Db        257 ThrAspGlyLeuLysIleSerAlaArgLeuAsnLeuLeuAspGlnGluLeuSerGluLys 276

Qy       1611 AGCGGGCAGCTGCGGCAGGGCAGCGCCCAGAGCCAGCAGCAGATCCGCGGGGAGATCGAC 1670
```

```
Db        277 SerLeuGlnLeuGluSerSerProThrGluGluLysMetLysIleSerGluGlnValGln 296

Qy       1671 AGCCTGCGCCAGGAGAAGGACTCGCTGCTCAAGCAGCGCCTGGAGATCGACGGCAAGCTG 1730
Db        297 AlaLeuGlnArgGluArgAspGlnLeuGlnArgGlnArgAsnSerValAspGluArgLeu 316

Qy       1731 AGGCAGGGGAGTCTGCTGTCCCCCGAGGAGGAGCGGACGCTGTTCCAGTTGGATGAGGCC 1790
Db        317 LysHisGlyArgValLeuSerProLysGluGluHisLeuLeuPheGlnLeuGluGluGly 336

Qy       1791 ATCGAGGCCCTGGATGCTGCCATTGAGTATAAGAATGAGGCCATCACATGCCGCCAGCGG 1850
Db        337 IleGluAlaLeuGluAlaAlaIleGluPheLysAsnGluSerIleGlnSerArgGlnAsn 356

Qy       1851 GTGCTTCGGGCCTCAGCCTCGTTGCTGTCCCAGTGCGAGATGAACCTCATGGCCAAGCTC 1910
Db        357 SerLeuLysAlaSerPheGlnAsnLeuSerGlnSerGluAlaAsnValLeuGluLysLeu 376

Qy       1911 AGCTACCTCTCATCCTCAGAGACCAGAGCCCTCCTCTGCAAGTATTTTGACAAGGTGGTG 1970
Db        377 ValCysLeuAsnIleThrGluIleArgAlaIleLeuPheLysTyrPheAsnLysValIle 396

Qy       1971 ACGCTCCGAGAGGAGCAGCACCAGCAGCAGATTGCCTTCTCGGAACTGGAGATGCAGCTG 2030
Db        397 AsnLeuArgGluThrGluArgLysGlnGlnLeuGlnAsnLysGluMetLysMetLysVal 416

Qy       2031 GAGGAGCAGCAGAGGCTGGTGTACTGGCTGGAGGTGGCCCTGGAGCGGCAGCGCCTGGAG 2090
Db        417 LeuGluArgAspAsnValValHisGluLeuGluSerAlaLeuGluHisLeuArgLeuGln 436

Qy       2091 ATGGACCGCCAGCTGACCCTGCAGCAGAAGGAGCACGAGCAGAACATGCAGCTGCTCCTG 2150
Db        437 CysAspArgArgLeuThrLeuGlnGlnLysGluHisGluGlnLysMetGlnLeuLeuLeu 456

Qy       2151 CAGCAGAGTCGAGACCACCTCGGTGAAGGGTTAGCAGACAGCAGGAGGCAGTATGAGGCC 2210
Db        457 GlnHisPheLysAspGlnAspGlyAspSerIleIleGluThrLeuLysAsnTyrGluAsp 476

Qy       2211 CGGATTCAAGCTCTGGAGAAGGAACTGGGCCGTTACATGTGGATAAACCAGGAACTGAAA 2270
Db        477 LysIleGlnGlnLeuGluLysAspLeuTyrPheTyrLysLysThrSerArgAspLeuLys 496

Qy       2271 CAGAAGCTC--------------GGCGGTGTG-----------AACGCTGTAGGCCAC 2303
Db        497 LysArgLeuLysAspProAlaGlnGlyAlaAlaGlnTrpGlnArgThrLeuThrGluHis 516

Qy       2304 AGCAGGGGTGGGGAG---------------------------------------------- 2318
Db        517 HisAspAlaGlyAspGlyValLeuAsnProGluGluThrThrValLeuSerGluGluLeu 536

Qy       2319 AAGAGGAGCCTGTGCTCGGAGGGCAGACAGGCTCCTGGAAATGAAGATGAGCTCCACCTG 2378
Db        537 LysTrpAlaSerArgThrGluAsnThrLysLeuAsnGlySerGluArgGluValAspAsn 556

Qy       2379 GCACCCGAGCTTCTCTGGCTGTCCCCCCTCACTGAGGGGGCCCCCCGCACCCGGGAGGAG 2438
Db        557 SerSerSerSerLeuLysThrGlnProLeuThrGlnGlnIlePro---------GluAsp 573

Qy       2439 ACGCGGGACTTGGTCCACGCTCCGTTACCCTTGACCTGGAAACGCTCGAGCCTGTGTGGT 2498
Db        574 GlyProAspSerLeuProAla--------------------ArgSerSerIle---Ala 585

Qy       2499 GAGGAGCAGGGGTCCCCCGAGGAACTGAGGCAGCGGGAGGCGGCTGAGCCCCTGGTGGGG 2558
Db        586 ProSerSerGlyGlnLeuGlnSerIleAlaAspLysThrGluAlaArgProPheThrHis 605

Qy       2559 CGGGTGCTTCCTGTGGGTGAGGCAGGCCTGCCC---TGGAACTTTGGGCCTTTGTCC--- 2612
Db        606 SerGlnSerProValProHisGlnPheGlnProValArgSerIleGlyProLeuGlnGly 625

Qy       2613 ---AAGCCC------------CGGCGGGAACTGCGACGAGCCAGCCCGGGGATGATTGAT 2657
```

```
Db        626 ValLysProValLysLeuCysArgArgGluLeuArgGlnIleSerAlaMetGluLeuSer 645

Qy       2658 GTCCGGAAAAACCCCCTGTAAGCCCTCGGGGCAGACCCTGCCTTGGAGGGAGACTCCGAG 2717
              :::|||:::      |||   |||   |||               :::         :::
Db        646 LeuArgArgCysSerLeuGluAlaGlyGlyArgSerMetThrAlaAspSerLeuGluAsp 665

Qy       2718 CCTGCTGAA 2726
              |||   |||
Db        666 ProGluGlu 668

RESULT 4
Q95JC9
ID   Q95JC9      PRELIMINARY;      PRT;   676 AA.
AC   Q95JC9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Basic proline-rich protein.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Parotid gland;
RA   Zhang Q., Szalay A.A., Kyeyune-Nyombi E., Sands J.F., Oberg K.C.,
RA   Tieche J.-M., Leonora J.;
RT   "Cloning and expression of a novel proline-rich protein from porcine
RT   parotid glands.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY035849; AAK61383.1; -.
DR   InterPro; IPR005404; KV3.3_channel.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01582; KV33CHANNEL.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   676 AA;  62297 MW;  3008BC41EFD81FC9 CRC64;

Alignment Scores:
Pred. No.:              1.9e-21       Length:        676
Score:                  626.50        Matches:       258
Percent Similarity:     29.54%        Conservative:  12
Best Local Similarity:  28.23%        Mismatches:    331
Query Match:            10.96%        Indels:        313
DB:                     6             Gaps:          47

US-10-032-996-6 (1-3121) x Q95JC9 (1-676)

Qy       2706 CCTCCAAGGCAGGGTCTGCCCCGAGGGCTTACAGGGGGTTTTTCCGGACATCAATCATCC 2647
              |||||||||        |||          |||        |||||||||
Db         43 ProProArgProProProProGluGluSerGlnGly------GluGlyHisGlnLysArg 60

Qy       2646 CCGGGCTGGCTCGTCGCAGTTCCCGCCGGGGCTTGGACAAAGGCCCAAAGTTCCAGGGCA 2587
              |||                  |||   |||
Db         61 ProArg--------------ProProGly------------------------------ 65

Qy       2586 GGCCTGCCTCACCCACAGGAAGCACCCGCCCCACCAGGGGCTCAGCCGCCTCCCGCTGCC 2527
                    |||:::::   ||||||||||||||||||:::|||||||||
Db         66 ------AspGlyProGluGlnGlyProAlaProProGlyAlaGluProProProGlyPro 83

Qy       2526 TCAGTTCCTCGGGGGACCCCTGCTCCTCACCACACAGGCTCGAGCGTTTCCAGGTCAAGG 2467
                 |||   |||   |||   |||          |||   :::         :::|||
Db         84 Pro---ProProGlyProProProPro---------GlyProAlaProProGlyAlaArg 99

Qy       2466 GTAACGGAGCGTGGACCAAGTCCCGCGTCTCCTCCCGGGTGCGGGGGGCCCCCTCAGTGA 2407
                       |||   |||     ||||||       |||||||||
Db        100 --------------ProProProGlyProProPro------ProGlyProProProPro 112

Qy       2406 GGGGGGACAGCCAGAGAAGCTCGGGTGCCAGGTGGAGCTCATCTTCATTTCCAGGAGCCT 2347
              |||          :::||||||   |||   |||
Db        113 GlyProAlaProProAspAlaArgProProProGly---------------------- 124

Qy       2346 GTCTGCCCTCCGAGCACAGGCTCCTCTTCTCCCCACCCCTGCTGTGGCCTACAGCGTTCA 2287
                    ||||||      |||         |||   |||
```

```
Db        125 ------ProProProProGlyProProProProGlyPro---------------------- 135

Qy       2286 CACCGCCGAGCTTCTGTTTCAGTTCCTGGTTTATCCACATGTAACGGCCCAGTTCCTTCT 2227
                              :::   ||||||              ||||||   |||
Db        136 -----------------AlaProProGlyAlaArgProProProGlyProProPro--- 148

Qy       2226 CCAGAGCTTGAATCCGGGCCTCATACTGCCTCCTGCTGTCTGCTAACCCTTCACCGAGGT 2167
                       ||||||
Db        149 ------------ProGlyPro----------------------------------------- 151

Qy       2166 GGTCTCGACTCTGCTGCAGGAGCAGCTGCATGTTCTGCTCGTGCTCCTTCTGCTGCAGGG 2107
                              |||   |||         ||||||   |||
Db        152 ---------ProProProGlyProAlaProProGlyAlaArgProProProGlyProPro 168

Qy       2106 TCAGCTGGCGGTCCATCTCCAGGCGCTGCCGCTCCAGGG---------------------- 2068
                 |||   |||   ||||||   |||   ||||||
Db        169 ProProGlyProProProProGlyProAlaProProGlyAlaArgProProProGlyPro 188

Qy       2067 CCACCTCCAGCCAGTACACCAGCCTCTGCTGCTCCTCCAGCTGCATCTCCAGTTCCGAGA 2008
              |||||||||        |||        |||||||||   |||   |||   |||
Db        189 ProProProGlyProProProProGlyProAlaProProGlyAlaArgProProPro--- 207

Qy       2007 AGGCAATCTGCTGCTGGTGCTGCTCCTCTCGGAGCGTCACCACCTTGTCAAAATACTTGC 1948
                              |||   |||   |||       ||||||
Db        208 --------------GlyProProProProGlyProProProPro--------------- 217

Qy       1947 AGAGGAGGGCTCTGGTCTCTGAGGATGAGAGGTAGCTGAGCTTGGCCATGAGGTTCATCT 1888
                                                                  |||   :::
Db        218 ----------------------------------------------------GlyProAla 220

Qy       1887 CGCACTGGGACAGCAACGAGGCTGAGGCCCGAAGCACCCGCTGGCGGCATGTGATGGCCT 1828
                    |||         ||||||
Db        221 ProProGlyAla-----------ArgProPro--------------------------- 227

Qy       1827 CATTCTTATACTCAATGGCAGCATCCAGGGCCTCGATGGCCTCATCCAACTGGAACAGCG 1768
                                    |||||||||   |||   |||
Db        228 ----------------------ProGlyProProProProGlyProProProProGly 239

Qy       1767 TCCGCTCCTCCTCGGGGGACAGCAGACTCCCCTGCCTCAGCTTGCCGTCGATCTCCAGGC 1708
                |||||||||                                  |||   ||||||
Db        240 ProAlaProPro--------------------------GlyAlaArgProProProGly 250

Qy       1707 GCTGCTTGAGCAGCGAGTCCTTCTCCTGGC---------GCAGGCTGTCGATCTCCCCGC 1657
                          |||   ||||||              |||   |||   |||
Db        251 ProProProLeuGlyProProProProGlyProAlaProProGlyAlaArgProProPro 270

Qy       1656 GGATCTGCTGCTGGCTCTGGGCGCTGCCCTGCCGCAGCT---------GCCCGCTCTTCT 1606
                |||                        |||       |||         ||||||
Db        271 GlyProProProProGlyProProProProGlyProAlaProProGlyAlaArgProPro 290

Qy       1605 CGGACAGCTCCTTCTCCAGGTGCTCCAGCCGGCTGGACACTCGCACGATGTCCTCGTTGA 1546
                   |||   ||||||   |||
Db        291 ProGlyProProProProGlyProProPro----------------------------- 300

Qy       1545 GGGCCTGGCTGGATCTCAGGCGCTTGCTCTCCAGCCCCGTCTTCTCCTGCATCAGGGCCT 1486
                 ||||||                  |||   |||       |||       ||||||
Db        301 ---ProGlyProAlaProProGlyAlaArgProProProGlyProProProProGlyPro 319

Qy       1485 ------------CCTTCTTGGCCAGGATGGCCTCCCGCTTGTGGAGCTCCTCCCCCAGCT 1438
                       |||:::   ||||||      |||      |||   |||||||||
Db        320 ProProProGlyProAlaProProGlyAlaArgProProProGlyProProProGly 339

Qy       1437 CCTCCAGCGCCCGCC---GCTGCTGTAGCACCTTCTCCATCTCCTGGTCCAGCCACTTCT 1381
                 ||||||   |||      |||                 |||   |||||||||
Db        340 ProProProProGlyProAlaProProGlyAlaArgProProProGlyProProProPro 359

Qy       1380 TCTGCTCCTCAATCTTCTGCTGCTGTTCCAGGCTGACCACAGAGCCGTTGCTGCCACTGC 1321
                    |||           |||   ||||||
Db        360 GlyProProProProGlyProAlaProProGly--------------------------- 370

Qy       1320 GCCTCTTCCTCTGGAAGGCCGCGATCTCTTCCGTCTTAATCTTCAGGATCTTCTGCTGTT 1261
                          ||||||      |||
Db        371 ------------AlaArgProProProGlyPro------------------------- 377
```

```
Qy       1260 GCTCATGCTTCAGCTCCAGCTCCTTGACGCGGTGCTGCCGCTTGCTCATTTCTGCCTCCA 1201
Db        378 --------------ProProPro------GlyProProProProGlyProAlaProPro 390

Qy       1200 GGCGCCGCTTCTGCTCCGTCTCCTCGCGAAGCCGCCTCTGCAGCTGTCCCTGCTGCTGCC 1141
Db        391 GlyAlaArgProProProGlyPro---ProProProGlyProProProProGlyProAla 409

Qy       1140 GCATGAGCTGCACGT---------TCCGCTCGAGCTCCTGCAGTCGCTTCTCACTCTGGG 1090
Db        410 ProProGlyAlaArgProProProProProProProProAlaAspGluProGlnGlnGly 429

Qy       1089 CCGACAGTGACACCAGCCGCTCCGTAGCCTGCTTCTTCTCCTTCAGCACCTGCA------ 1036
Db        430 Pro------AlaProSerGlyAspLysProLysLysLysProProProProAlaGlyPro 447

Qy       1035 CCTGGCTCTGGGCCGCAGCGACCCTCCTGCGGAACTCCTGGAGCCGAGACCGCTCGCCAG 976
Db        448 ProProProGlyProProSerProGlyProAlaProProGly---------AlaArgPro 464

Qy        975 CATCCTGGAGCTCCTTGCCCTCGAGCTCCCGCAGCTGCCTCTGGCCTTCACTCAGCTCGG 916
Db        465 ProProGlyProProProProGlyProProProProGlyProAlaProProGlyAlaArg 484

Qy        915 CCCGCACCTGCTCTG------------CCTCCTGCTCCAGCTCCCGGATACGCTGGCTGT 868
Db        485 ProProProGlyProProProProGlyProProProProGlyProAlaProProGlyAla 504

Qy        867 GCTGGCGGTTCAGGGCCTGAGCTGCCTTTCCTGTGCGGACCAGCTCGCCAATAAGCTCCT 808
Db        505 ArgProProProGlyProProProPro--------------------------GlyPro 515

Qy        807 CCTTCATGCGGATGTTGATAGCCAGCT---CCCGGATCTTCTGCTGGGCCTGGGCCAGCC 751
Db        516 ProProProGly---------ProAlaProProGlyAlaArgProProProGlyProPro 532

Qy        750 GCCACTCTGAGGCTGTGGCAGGGGGGACCTGGCGGGCCTGAACTCGGGCCTTGCTCCCAC 691
Db        533 Pro------------------------ProGlyProProProProGlyProAlaPro--- 543

Qy        690 CAACTGCTCTGGACCCTGGAATGGCTGCATCCAACTCCTCAAGGCAAAGCTCTGGGCCCT 631
Db        544 ------------------ProGlyAlaArgProProProGlyProProProProGlyPro 557

Qy        630 TCCTCTCTGGCAGACTCCCTGGGCGTGCCCCCGCCCTCTGACTGCAGTTGCTGATCCTAT 571
Db        558 --------------------------ProProProGlyProAlaPro------------ 564

Qy        570 TTCTGCGCAGGTGTAAGGTCCGCCTGGGCGGCTCCTCCTCCTCCTCTTCCTCCTCTGAAG 511
Db        565 ------ProGlyAlaArgProPro---ProGlyProProProProGlyProPro------ 579

Qy        510 CAGCTGAAGAGCCACTTCCCAGCCTGTTCACCTCAGTCAGCAACTCAGCTCCAGCCTCCC 451
Db        580 ------------------------------------------------------ProPro 581

Qy        450 TGCCATTTGTCACCTGCTCTCCCCTCTGCTCAGAGCCAACTTCATCTCCAGGGAGGCAGG 391
Db        582 ---------GlyProAlaProProGlyAlaArg-------------------------- 589

Qy        390 CAGGCGGCACCATGCCCAGCACATGGGCGTGGGCACCCCCCAGGGGGGCTGTATGAGGTC 331
Db        590 ---------ProProPro------------GlyProProPro----------------- 596

Qy        330 GAGGCACAAAGGACCCGGGAGGCAGGCCATTCAGGAGCCGCAGGCCCCCCCAGCCTGGCC 271
Db        597 ------------------------------------ProGlyProProPro---Pro 602

Qy        270 GCACCAGCTCTAACCGCAGCCGCAGTTCCACCATCTCCTCCTGCTGCTCACGCAGCCGGT 211
Db        603 GlyProAlaProProGlyAlaArgProProProGlyProPro---------ProProGly 619
```

```
Qy        210 CGCTCTGCAGTTTGTACTGCTCCA------TGGCGTCCTCCAGCGCAGCCAGAAAGTCTC 157
Db        620 ProProProProGlyProAlaProProGlyAlaArgProProProGlyProProProPro 639

Qy        156 GGTTCTCCTCCTCCAGCCGCGCCACCTGGTTCTGCAGGGTCAGCAGCTGCTGCGCCCCCT 97
Db        640 GlyProProProProGlyProAlaProProGlyAlaArgProProProGlyProProPro 659

Qy         96 CATCCTCCTTTCGCCCGCCGGCCCCCTGCGCCGCCTGGTCCT 55
Db        660 ProProProGlyProSerProProArgProProProGlyPro 673
```

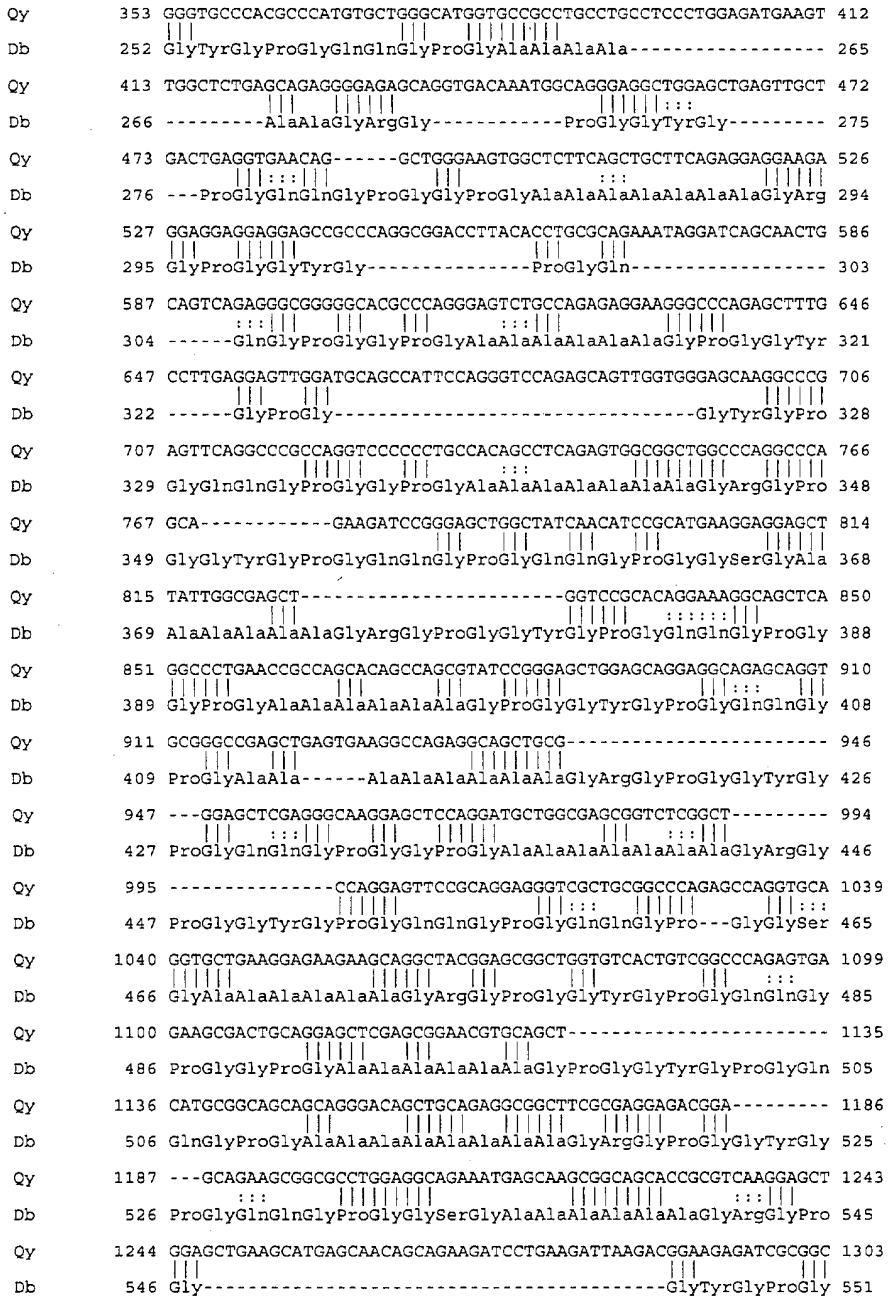
RESULT 5
Q9BIT7
ID Q9BIT7 PRELIMINARY; PRT; 1953 AA.
AC Q9BIT7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Major ampullate spidroin 2-like protein (Fragment).
OS Nephila madagascariensis.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX NCBI_TaxID=115969;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
RT Fibroin Sequences.";
RL Science 291:2603-2605(2001).
DR EMBL; AF350276; AAK30605.1; -.
FT NON_TER 1 1
FT NON_TER 1953 1953
SQ SEQUENCE 1953 AA; 159383 MW; C82B4DACD043C3BD CRC64;

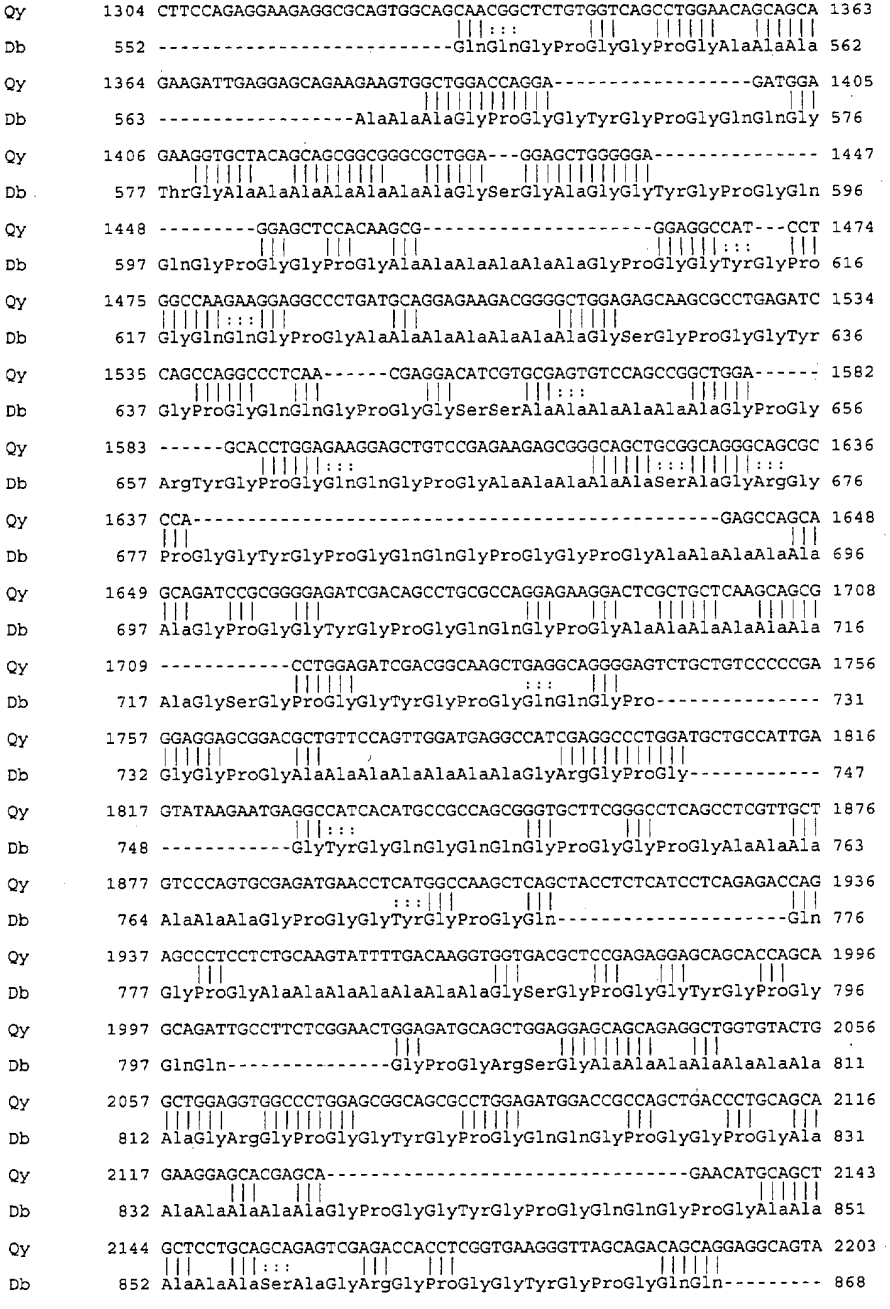
Alignment Scores:

Pred. No.:	4.09e-20	Length:	1953
Score:	598.00	Matches:	309
Percent Similarity:	32.35%	Conservative:	41
Best Local Similarity:	28.56%	Mismatches:	409
Query Match:	10.42%	Indels:	323
DB:	5	Gaps:	51

US-10-032-996-6 (1-3121) x Q9BIT7 (1-1953)

```
Qy         53 CGAGGACCAGGCGGCGCAGGG---GGCCGGCGGGCGAAAGGAGGATGAGGGGGCGCAGCA 109
Db        120 ArgGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyProGlyAlaAlaAla 139

Qy        110 GCTGCT----------------------------------GACCCTGCAGAACCAGGTGGC 136
Db        140 AlaAlaAlaGlySerGluGlyTyrGlyProGlyGlnGlnGlyProArgGlyProGlyAla 159

Qy        137 GCGGCTGGAGGA--------------------------GGAGAACCGAGACTTTCTGGC 169
Db        160 AlaAlaAlaGlyProGlyGlyTyrGlyProGlyGlnGlnGlyAlaSerAlaAlaAlaSer 179

Qy        170 TGCGCTGGAGGACGC------------CATGGAGCAGTACAAACTGCAGAGCGACCGGCT 217
Db        180 AlaAlaAlaGlyArgGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyPro 199

Qy        218 GCGTGAGCAGCAGGAGGAGATGGTGGA--------------------------ACTGCG 250
Db        200 SerAlaAlaAlaAlaGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProSerAlaAla 219

Qy        251 GCTGCGGTTAGAGCTGGTGCGGCCAGGCTGGGG-----------------GGGCCTGCG 292
Db        220 AlaAlaAlaAlaAlaGlySerGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGly 239

Qy        293 GCTCCTGAATGGCCTGCCTCCCGGGTCCTTTGTGCCTCGACCTCATACAGCCCCCCTGGG 352
Db        240 GlyProGly----------------------AlaAlaAlaAlaAlaAlaGlyProGly 251
```

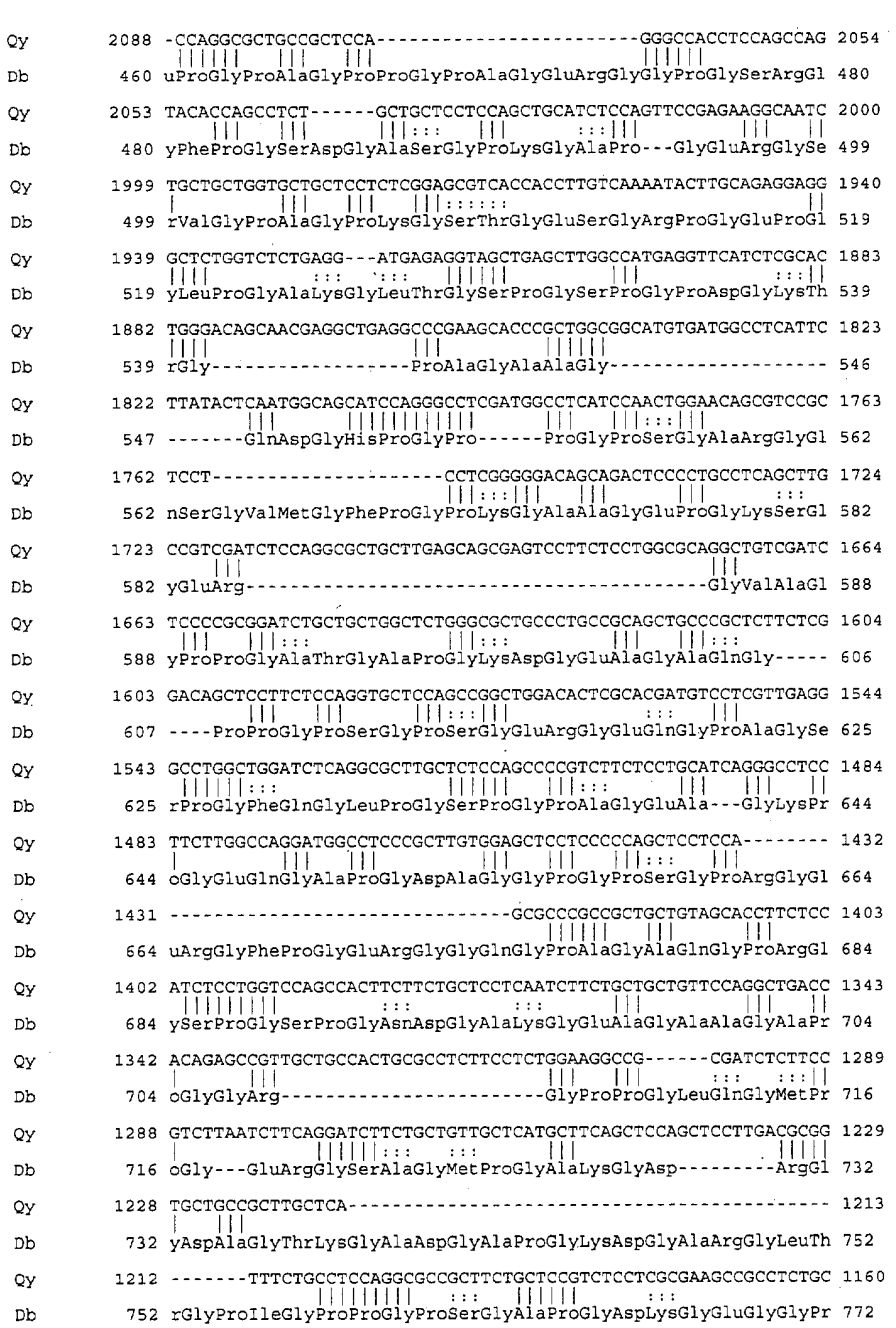

Qy 353 GGGTGCCCACGCCCATGTGCTGGGCATGGTGCCGCCTGCCTGCCTCCCTGGAGATGAAGT 412
Db 252 GlyTyrGlyProGlyGlnGlnGlyProGlyAlaAlaAlaAla---------- 265
Qy 413 TGGCTCTGAGCAGAGGGGAGAGCAGGTGACAAATGGCAGGGAGGCTGGAGCTGAGTTGCT 472
Db 266 ---------AlaAlaGlyArgGly------------ProGlyGlyTyrGly--------- 275
Qy 473 GACTGAGGTGAACAG------GCTGGGAAGTGGCTCTTCAGCTGCTTCAGAGGAGGAAGA 526
Db 276 ---ProGlyGlnGlnGlyProGlyGlyProGlyAlaAlaAlaAlaAlaAlaGlyArg 294
Qy 527 GGAGGAGGAGGAGCCGCCCAGGCGGACCTTACACCTGCGCAGAAATAGGATCAGCAACTG 586
Db 295 GlyProGlyGlyTyrGly---------------ProGlyGln----------------- 303
Qy 587 CAGTCAGAGGGCGGGGGCACGCCCAGGGAGTCTGCCAGAGAGGAAGGGCCCAGAGCTTTG 646
Db 304 ------GlnGlyProGlyGlyProGlyAlaAlaAlaAlaAlaAlaGlyProGlyGlyTyr 321
Qy 647 CCTTGAGGAGTTGGATGCAGCCATTCCAGGGTCCAGAGCAGTTGGTGGGAGCAAGGCCCG 706
Db 322 ------GlyProGly---------------------------------GlyTyrGlyPro 328
Qy 707 AGTTCAGGCCCGCCAGGTCCCCCCTGCCACAGCCTCAGAGTGGCGGCTGGCCCAGGCCCA 766
Db 329 GlyGlnGlnGlyProGlyGlyProGlyAlaAlaAlaAlaAlaAlaAlaGlyArgGlyPro 348
Qy 767 GCA----------GAAGATCCGGGAGCTGGCTATCAACATCCGCATGAAGGAGGAGCT 814
Db 349 GlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlnGlnGlyProGlySerGlyAla 368
Qy 815 TATTGGCGAGCT----------GGTCCGCACAGGAAAGGCAGCTCA 850
Db 369 AlaAlaAlaAlaAlaGlyArgGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGly 388
Qy 851 GGCCCTGAACCGCCAGCACAGCCAGCGTATCCGGGAGCTGGAGCAGGAGGCAGAGCAGGT 910
Db 389 GlyProGlyAlaAlaAlaAlaAlaGlyProGlyGlyTyrGlyProGlyGlnGlnGly 408
Qy 911 GCGGGCCGAGCTGAGTGAAGGCCAGAGGCAGCTGCG---------- 946
Db 409 ProGlyAlaAla------AlaAlaAlaAlaAlaAlaGlyArgGlyProGlyGlyTyrGly 426
Qy 947 ---GGAGCTCGAGGGCAAGGAGCTCCAGGATGCTGGCGAGCGGTCTCGGCT--------- 994
Db 427 ProGlyGlnGlnGlyProGlyGlyProGlyAlaAlaAlaAlaAlaAlaGlyArgGly 446
Qy 995 ----------CCAGGAGTTCCGCAGGAGGGTCGCTGCGGCCCAGAGCCAGGTGCA 1039
Db 447 ProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlnGlnGlyPro---GlyGlySer 465
Qy 1040 GGTGCTGAAGGAGAAGAAGCAGGCTACGGAGCGGCTGGTGTCACTGTCGGCCCAGAGTGA 1099
Db 466 GlyAlaAlaAlaAlaAlaGlyArgGlyProGlyGlyTyrGlyProGlyGlnGlnGly 485
Qy 1100 GAAGCGACTGCAGGAGCTCGAGCGGAACGTGCAGCT---------- 1135
Db 486 ProGlyGlyProGlyAlaAlaAlaAlaAlaGlyProGlyGlyTyrGlyProGlyGln 505
Qy 1136 CATGCGGCAGCAGCAGGGACAGCTGCAGAGGCGGCTTCGCGAGGAGACGGA--------- 1186
Db 506 GlnGlyProGlyAlaAlaAlaAlaAlaAlaAlaGlyArgGlyProGlyGlyTyrGly 525
Qy 1187 ---GCAGAAGCGGCGCCTGGAGGCAGAAATGAGCAAGCGGCAGCACCGCGTCAAGGAGCT 1243
Db 526 ProGlyGlnGlnGlyProGlyGlySerGlyAlaAlaAlaAlaAlaGlyArgGlyPro 545
Qy 1244 GGAGCTGAAGCATGAGCAACAGCAGAAGATCCTGAAGATTAAGACGGAAGAGATCGCGGC 1303
Db 546 Gly----------GlyTyrGlyProGly 551

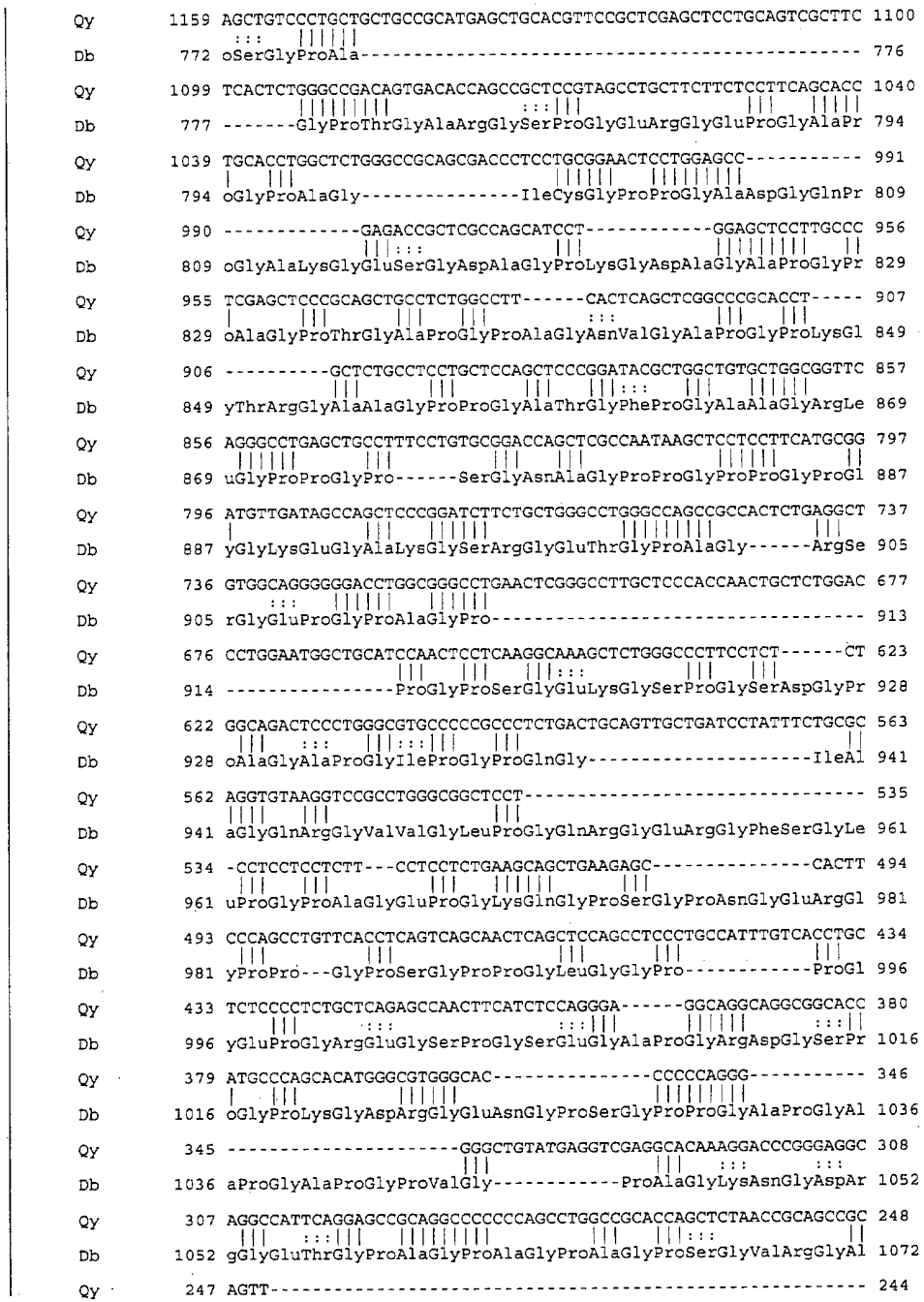

Qy 1304 CTTCCAGAGGAAGAGGCGCAGTGGCAGCAACGGCTCTGTGGTCAGCCTGGAACAGCAGCA 1363
Db 552 ----------GlnGlnGlyProGlyGlyProGlyAlaAlaAla 562
Qy 1364 GAAGATTGAGGAGCAGAAGAAGTGGCTGGACCAGGA----------GATGGA 1405
Db 563 ----------AlaAlaAlaGlyProGlyGlyTyrGlyProGlyGlnGlnGly 576
Qy 1406 GAAGGTGCTACAGCAGCGGCGGGCGCTGGA---GGAGCTGGGGGA---------- 1447
Db 577 ThrGlyAlaAlaAlaAlaAlaAlaGlySerGlyAlaGlyGlyTyrGlyProGlyGln 596
Qy 1448 ---------GGAGCTCCACAAGCG----------GGAGGCCAT---CCT 1474
Db 597 GlnGlyProGlyGlyProGlyAlaAlaAlaAlaAlaGlyProGlyGlyTyrGlyPro 616
Qy 1475 GGCCAAGAAGGAGGCCCTGATGCAGGAGAAGACGGGGCTGGAGAGCAAGCGCCTGAGATC 1534
Db 617 GlyGlnGlnGlyProGlyAlaAlaAlaAlaAlaAlaGlySerGlyProGlyGlyTyr 636
Qy 1535 CAGCCAGGCCCTCAA------CGAGGACATCGTGCGAGTGTCCAGCCGGCTGGA------ 1582
Db 637 GlyProGlyGlnGlnGlyProGlyGlySerSerAlaAlaAlaAlaAlaGlyProGly 656
Qy 1583 ------GCACCTGGAGAAGGAGCTGTCCGAGAAGAGCGGGCAGCTGCGGCAGGGCAGCGC 1636
Db 657 ArgTyrGlyProGlyGlnGlnGlyProGlyAlaAlaAlaAlaAlaSerAlaGlyArgGly 676
Qy 1637 CCA----------GAGCCAGCA 1648
Db 677 ProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyProGlyAlaAlaAlaAlaAla 696
Qy 1649 GCAGATCCGCGGGGAGATCGACAGCCTGCGCCAGGAGAAGGACTCGCTGCTCAAGCAGCG 1708
Db 697 AlaGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyAlaAlaAlaAlaAlaAla 716
Qy 1709 ----------CCTGGAGATCGACGGCAAGCTGAGGCAGGGGAGTCTGCTGTCCCCCGA 1756
Db 717 AlaGlySerGlyProGlyGlyTyrGlyProGlyGlnGlnGlyPro---------- 731
Qy 1757 GGAGGAGCGGACGCTGTTCCAGTTGGATGAGGCCATCGAGGCCCTGGATGCTGCCATTGA 1816
Db 732 GlyGlyProGlyAlaAlaAlaAlaAlaAlaGlyArgGlyProGly---------- 747
Qy 1817 GTATAAGAATGAGGCCATCACATGCCGCCAGCGGGTGCTTCGGGCCTCAGCCTCGTTGCT 1876
Db 748 ----------GlyTyrGlyGlnGlyGlnGlyProGlyProGlyAlaAlaAla 763
Qy 1877 GTCCCAGTGCGAGATGAACCTCATGGCCAAGCTCAGCTACCTCTCATCCTCAGAGACCAG 1936
Db 764 AlaAlaAlaGlyProGlyGlyTyrGlyProGlyGln----------Gln 776
Qy 1937 AGCCCTCCTCTGCAAGTATTTTGACAAGGTGGTGACGCTCCGAGAGGAGCAGCACCAGCA 1996
Db 777 GlyProGlyAlaAlaAlaAlaAlaAlaGlySerGlyProGlyGlyTyrGlyProGly 796
Qy 1997 GCAGATTGCCTTCTCGGAACTGGAGATGCAGCTGGAGGAGCAGCAGAGGCTGGTGTACTG 2056
Db 797 GlnGln----------GlyProGlyArgSerGlyAlaAlaAlaAlaAlaAlaAla 811
Qy 2057 GCTGGAGGTGGCCCTGGAGCGGCAGCGCCTGGAGATGGACCGCCAGCTGACCCTGCAGCA 2116
Db 812 AlaGlyArgGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyProGlyAla 831
Qy 2117 GAAGGAGCACGAGCA----------GAACATGCAGCT 2143
Db 832 AlaAlaAlaAlaAlaGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyAlaAla 851
Qy 2144 GCTCCTGCAGCAGAGTCGAGACCACCTCGGTGAAGGGTTAGCAGACAGCAGGAGGCAGTA 2203
Db 852 AlaAlaAlaSerAlaGlyArgGlyProGlyGlyTyrGlyProGlyGlnGln--------- 868
Qy 2204 TGAGGCCCGGATTCAAGC----------TCTGGAGAA 2230

```
Db        869 ---GlyProGlyGlySerGlyAlaAlaAlaAlaAlaGlyArgGlyProGlyGlyTyr 887

Qy       2231 GGAACTGGGCCGTTACATGTGGATAAACCAGGAACTGAAACAGAAGCTCGGCGGTGTGAA 2290

Db        888 GlyProGlyGlnGlnGlyProGlyGlyProGlyAla---AlaAlaAlaAlaAlaGly 906

Qy       2291 CGCTGTAGGCCACAGCAGGGGTGGGGAGAAGAGGAGCCTGTGCTCGGAGGGCAGACAGGC 2350

Db        907 ArgGlyProGlyGlyTyrGlyProGlyGlnGlnGlyPro---------GlyGlnGlnGly 923

Qy       2351 TCCTGGAAATGAAGATGAGCTCCACCTGGCACCCGAGCTTCTCTGGCTGTCCCC------ 2404

Db        924 --------------------ProGlySerGlyAlaAlaAlaAlaAlaGlyArg 936

Qy       2405 ------------------CCTCACTGAGGGGGCCCCCCGCACCCGGGA------------ 2434

Db        937 GlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyProGlyAlaAlaAlaAla 956

Qy       2435 ------GGAGACGCGGGACTTGGTCCACGCTCCGTTACCCTTGACCTGGAAACGCTCGAG 2488

Db        957 AlaAlaGlyProGlyGlyTyrGlyPro------------------------------- 965

Qy       2489 CCTGTGTGGTGAGGAGCAGGGGTCCCCCGAGGAACTGAGGCAGCGGGAGGCGGCTGAGCC 2548

Db        966 -----------GlyGlnGlnGlyProGlyAlaAlaAlaAlaAlaAlaGlySerGly 981

Qy       2549 CCTGGTGGGGCGGGT---GCTTCCTGTGGGTGAGGCAGGCCTGCCCTGGAACTTTGGGCC 2605

Db        982 ProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyProGlyAlaAlaAlaAlaAla 1001

Qy       2606 TTTGTCCAAGCCCCGGCGGGAACTGCGACGAGCCAGCCCGGGGATGATTGATGTCCGGAA 2665

Db       1002 AlaGlyArgGlyProGlyGlyTyrGlyProGlyGlnGlnGly------------------ 1015

Qy       2666 AAACCCCCTGTAAGCCCTCGGGGCAGACCCTGCCTTGGAGGGAGACTCCGAGCCTGCTGA 2725

Db       1016 ------ProGlyGlyPro-GlyAlaAlaAlaAlaAlaAlaAlaGlyArgGlyProGlyGl 1033

Qy       2726 AAGGGGCAGCTGCCTGTTTTGCTTCTGTGAAGGGCAGTCCTTACCGCACACCCTAAATCC 2785

Db       1033 yTyrGly---------------------ProGlyGlnGlnGlyProGlyGlnGlnGlyPr 1046

Qy       2786 AGGC 2789

Db       1046 oGly 1047
```

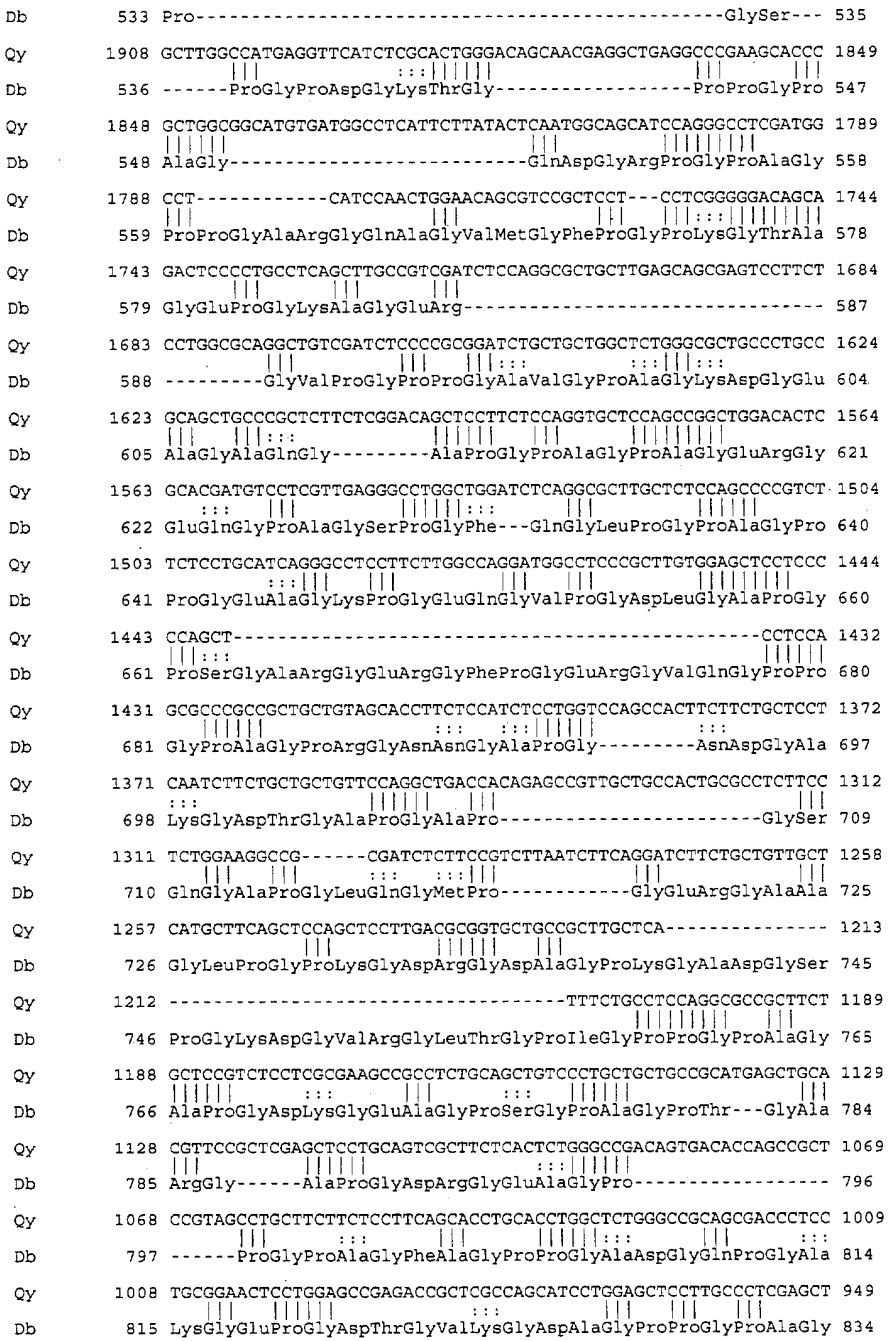
```
RESULT 6
Q9YIB4
ID   Q9YIB4      PRELIMINARY;      PRT;  1450 AA.
AC   Q9YIB4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Alpha 1 type I collagen.
OS   Cynops pyrrhogaster (Japanese common newt).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX   NCBI_TaxID=8330;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Regenerate forelimbs;
RX   MEDLINE=99407244; PubMed=10474166;
RA   Asahina K., Obara M., Yoshizato K.;
RT   "Expression of genes of type I and type II collagen in the formation
RT   and development of the blastema of regenerating newt limb.";
RL   Dev. Dyn. 216:59-71(1999).
DR   EMBL; AB015438; BAA36973.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Collagen; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Collagen.
SQ   SEQUENCE   1450 AA;  137564 MW;  ABF8A74841B87B7C CRC64;
```

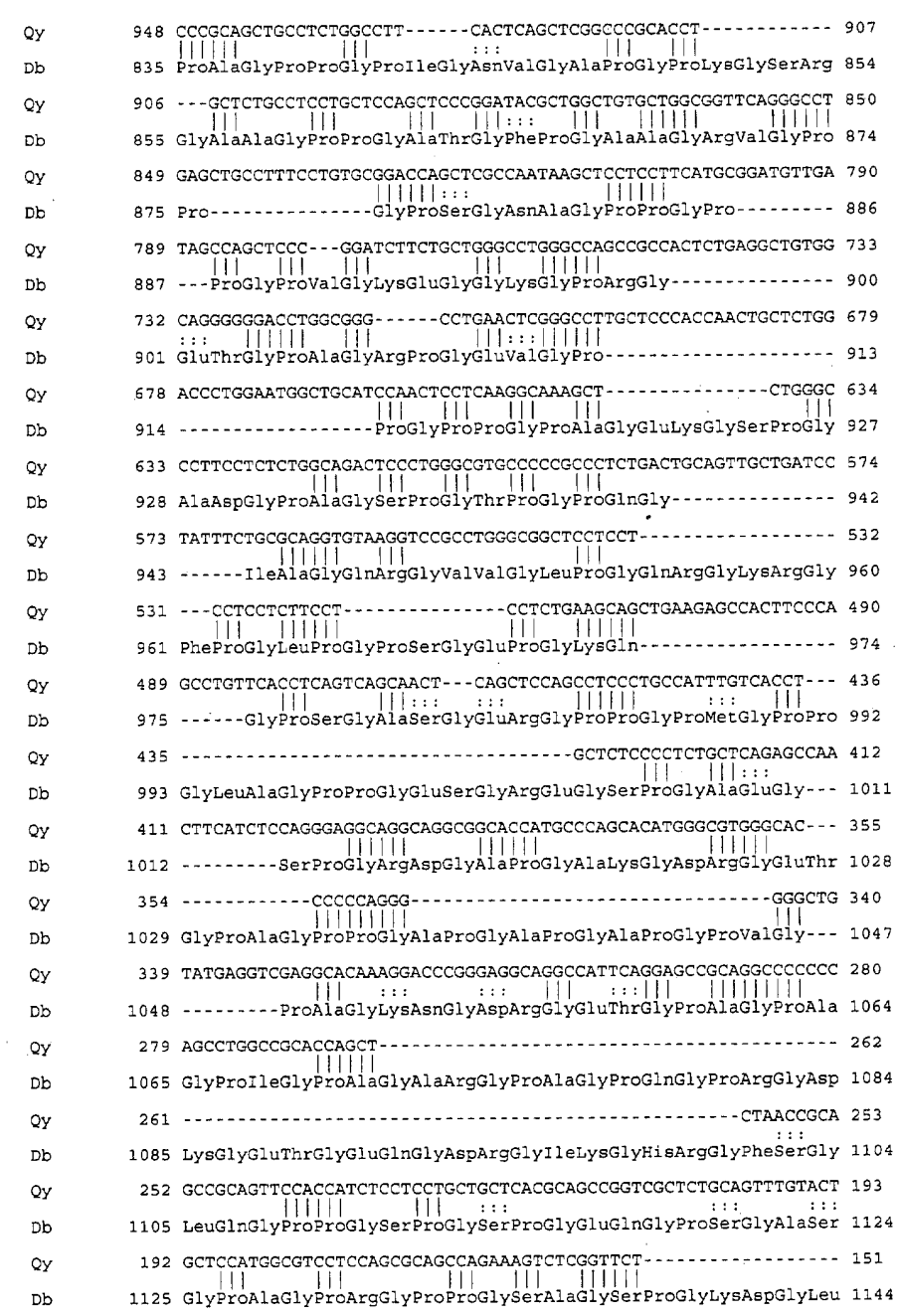
```
Alignment Scores:
Pred. No.:                2.19e-18        Length:          1450
Score:                    560.50          Matches:         291
Percent Similarity:       31.58%          Conservative:    63
Best Local Similarity:    25.96%          Mismatches:      416
Query Match:              9.81%           Indels:          351
DB:                       13              Gaps:            56

US-10-032-996-6 (1-3121) x Q9YIB4 (1-1450)

Qy       2770 GGTAAGGACTGCCCTTCACAGAAGCAAAACAGGCAGCTGCCCCTTTCAGCAGGCTCGGAG 2711

Db        216 GlyAspAspGlyGluSerGlyLysProGlyArg--------------------------- 226

Qy       2710 TCTCCCTCCAAGGCAGGG---TCTGCCCCGAGGGCTTACAGGGGGTTTTTCCGGACATCA 2654

Db        227 ---ProGlyGluArgGlyProSerGlyProGlnGlyAlaArgGlyLeuProGlyThrAla 245

Qy       2653 ATCATCCCCGGGCTGGCTCGTCGCAGTTCCCGCCGGGGCTTGGACAAAGGCCCAAAGTTC 2594

Db        246 GlyLeuProGlyMetLysGlyHisArgGlyPheAsnGlyLeuAsp---GlyAlaLysGly 264

Qy       2593 CAGGGCAGGCCTGCCTCACCCACAGGAAGCACCCGCCCCACCAGGGGCTCAGCCGCCTCC 2534

Db        265 AspAsnGlyProAlaGlyProLysGlyGlu-ProGlyAsnProGlyGluAsnGlyAlaPr 284

Qy       2533 CGCTGCCTCAGTTCCTCGGGGGACCCCT----------------------GCTCCTCACCA 2495

Db        284 oGlyGlnAlaGlyProArgGlyLeuProGlyGluArgGlyArgProGlyAlaProGlyPr 304

Qy       2494 CACAGGCTCGAGCGTTTCCAGGTCAAGGGTAACGGAGCGTGGACCAAGTCCCGCGTCTCC 2435

Db        304 oAlaGly---------------AlaArgGlyAsnAspGlySerProGlyAlaAlaGlyPr 319

Qy       2434 TCCCGGGTGCGGGGGGCCC---------------------------------------- 2416

Db        319 oProGlyProThrGlyProThrGlyProProGlyPheProGlyAlaValGlyAlaLysGl 339

Qy       2415 ----------CCTCAGTGAGGGGGGACAGCCAGAGAAGCTCGGGTGCCAGGTGGAGCTCA 2366

Db        339 yAspAlaGlyProGln---------GlySerArgGlySerGluGlyProGlnGlyAlaAr 356

Qy       2365 TCTTCATTTCCAGGAGCCTGTCTGCCCTCCGAGCACAGGCTCCTCTTCTCCC------CA 2312

Db        356 g---------GlyGluProGlyAlaProGlyProAlaGlyAlaAlaGlyProSerGlyAs 373

Qy       2311 CCCCTGCTGTGGCCTACAGCGTTCACACCGCCGAGCTTCTGTTTCA-------------- 2266

Db        373 nProGlyThrAspGlyGlnProGlyGlyLysGlyAlaThrGlySerProGlyIleAlaGl 393

Qy       2265 ----------GTTCCTGGTTTATCCACATGTAACGGCCCA--------------GTTCC 2231

Db        393 yAlaProGlyPheProGlyAlaArgGlyAlaProGlyProGlnGlyProAlaGlyAlaPr 413

Qy       2230 TTCTCCAGAGCTTGAATCCGGGCCTCATACTGCCTCCTGCTGTCTGCTAACCCTTCACCG 2171

Db        413 oGlyProLysGlyAsnAsnGlyGluProGlyAlaGln----------------------- 425

Qy       2170 AGGTGGTCTCGACTCTGCTGCAGGAGCAGCTGCATGTTCTGCTCGTGCTCCTTCTGCTGC 2111

Db        426 ----GlyAsnLysGlyGluProGlyAlaLysGly------------GluProGlyProAl 440

Qy       2110 AGGGTCAGCTGGC------GGTCCATCT------------------------------ 2089

Db        440 aGlyValGlnGlyProProGlyProSerGlyGluGluGlyLysArgGlySerArgGlyGl 460
```





```
Db       1072 aProGlyProAlaGlyAlaArgGlyAspLysGlyGluAlaGlyGluGlnGlyGluArgGl 1092

Qy        243 ------------------------------------CCACCATCTCCTCCTGCTGCTCA 221

Db       1092 yMetLysGlyHisArgGlyPheAsnGlyMetGlnGlyProProGlyProProGlySerSe 1112

Qy        220 CGCAGCCGGTCGCTCTGCAGTTTGTACTGCTCCATGGCGTCCTCCAGCGCAGCCAGAAAG 161

Db       1112 rGlyGluGlnGlyAlaProGlyProSerGlyProAlaGlyProArgGlyProProGlySe 1132

Qy        160 TCTCGGTTCT--------------------------CCTCCTCCAGCCGCGCCACCTGG 128

Db       1132 rSerGlySerThrGlyLysAspGlyValAsnGlyLeuProGlyProIleGlyProProGl 1152

Qy        127 TTCTGCAGGGTCA------------GCAGCTGCTGCGCCCCCTCATCCTCCTTTCGCCCG 80

Db       1152 yProArgGlyArgAsnGlyAspValGlyProAlaGlyProProGlyProPro-------- 1169

Qy         79 CCGGCCCCCTGCGCCGCCTGGTCCTCGACGGAGGCGCTCTCGATGCCGCTATCGGGCCC 21

Db       1170 -GlyProProGlyProProGlyPro-ProSerGlyGlyPheAspPheSerPheMetPro 1188

RESULT 7
Q63079
ID   Q63079      PRELIMINARY;      PRT;  1453 AA.
AC   Q63079;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Collagen alpha1 (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1092 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Tooth;
RA   Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;
RT   "Expression of Collagen alpha1(I) mRNA variants during Tooth and Bone
RT   Formation in the Rat.";
RL   J. Dent. Res. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Tooth;
RA   Wurtz T.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z78279; CAB01633.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Collagen; 3.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Collagen.
FT   NON_TER       1      1
SQ   SEQUENCE   1453 AA;  137887 MW;  E6896BDC19A4A1D8 CRC64;

Alignment Scores:
Pred. No.:              2.19e-18        Length:         1453
Score:                  560.50          Matches:        295
Percent Similarity:     30.90%          Conservative:   57
Best Local Similarity:  25.90%          Mismatches:     420
Query Match:            9.81%           Indels:         367
DB:                     11              Gaps:           60

US-10-032-996-6 (1-3121) x Q63079 (1-1453)

Qy       2850 GGGTCCAGTTCTTTTGGGCCATGGCCCAAATTTGTTGATCCCAGTGAGGGTACAGATGAG 2791

Db        202 GlySerGlyProMetGlyProProGlyProProGlyProGlyLysAsnGlyAspAsp 221

Qy       2790 GGCCTGGATTTAGGGTGTGCGGTAAGGACTGCCCTTCACAGAAGCAAAACAGGCAGCTGC 2731

Db        222 Gly---GluAlaGlyLysProGlyArgProGly---GluArgGlyProProGlyPro--- 238

Qy       2730 CCCTTTCAGCAGGCTCGGAGTCTCCCTCCAAGGCAGGGTCTGCCC-------------- 2686

Db        239 ------GlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeuProGlyMetLysGlyHis 256

Qy       2685 CGAGGGCTTACAGGG----------------------GGTTTTTCCGGACATCAATCA 2650

Db        257 ArgGlyPheSerGlyLeuAspGlyAlaLysGlyAspThrGlyProAlaGlyProLysGly 276

Qy       2649 TCCCCGGGCTGGCTCGTCGCAGTTCCCGCCGGGGCTTGGACAAAGGCCCAAAGTTCCAGG 2590

Db        277 GluProGly------------SerProGlyGluAsnGlyThrProGlyGlnMetGlyPro 292

Qy       2589 GCAGGCCTGCCTCACCCACAGGAAGCACCCGCCCCACCA---------------------- 2551

Db        293 ArgGlyLeuProGlyGluArgGlyArgProGlyProProGlyThrAlaGlyAlaArgGly 312

Qy       2550 --------------GGGGCTCAGCCGCCTCCCGCTGCCTCAGTTCCTCGGGGGACCCCT 2506

Db        313 AsnAspGlyAlaValGlyAlaAlaGlyProProGlyProThrGlyProThrGlyProPro 332

Qy       2505 --------------------GCTCCTCACCACACAGGCTCGAGCGTTTCCAGGTCAAGG 2467

Db        333 GlyPheProGlyAlaAlaGlyAlaLysGlyGluAlaGlyProGlnGlyAlaArgGlySer 352

Qy       2466 GTAACGGAGCGTGGA------------CCAAGTCCCGCGTCTCCTCCCGGGTGCGGGGGG 2419

Db        353 GluGlyProGlnGlyValArgGlyGluProGlyProProGlyProAlaGlyAlaAlaGly 372

Qy       2418 CCCCCTCAGTGAGGGGGGACAGCCAGAGAA---------------------------- 2389

Db        373 ProAlaGlyAsnProGlyAlaAspGlyGlnProGlyAlaLysGlyAlaAsnGlyAlaPro 392

Qy       2388 ----------------------------GCTCGGGTGCCAGGTGGAGCTCATCTTCAT 2359

Db        393 GlyIleAlaGlyAlaProGlyPheProGlyAlaArgGlyProSerGlyPro--------- 409

Qy       2358 TTCCAGGAGCCTGTCTGCCCTCCGAGCACAGGCTCCTCTTCTCCCCACCCCTGCTGTGGC 2299

Db        410 ---GlnGlyProSerGlyAlaProGlyProLysGlyThrSerGlyGluProGlyAlaPro 428

Qy       2298 CTACAGCGTTCACACCGCCGAGCTTCTGTTTCAGTTCCTGGTTTATCCACATGTAACGGC 2239

Db        429 GlyAsnLysGlyAspThrGlyAlaLys---GlyGluProGlyProAlaGlyValGlnGly 447

Qy       2238 CCAGTTCCTTCTCCAGAGCTTGAATCCGGGCCTCATACTGCCTCCTGCTGTCTGCTAACC 2179

Db        448 Pro---ProGlyProAlaGlyGluGlyLysArgGlyAla------------------- 460

Qy       2178 CTTCACCGAGGT----------------GGTCTCGACTCTGCTGCAGGA------GCAGCT 2140

Db        461 ------ArgGlyGluProGlyProSerGlyLeuProGlyProProGlyGluArgGlyGly 478

Qy       2139 GCATGTTCTGCTCGTGCTCCTTCTGCTGCAGGGTCAGCTGGC------GGTCCATCTCCA 2086

Db        479 ProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyProLysGlyProSerGly 498

Qy       2085 GGCGCTGCCGCTCCAGGGCCACCT---CCAGCCAGTACACCAGCCTCTGCTGCTCCTCCA 2029

Db        499 GluArgGlySerProGlyProAlaGlyProLysGlySerProGlyGluAlaGlyArgPro 518

Qy       2028 GCTGCATCTCCAGTTCCGAGAAGGCAATCTGCTGCTGGTGCTGCTCCTCTCGGAGCGTCA 1969

Db        519 GlyGluAlaGlyLeuPro-----------------GlyAlaLysGlyLeuThrGlySer 532

Qy       1968 CCACCTTGTCAAAATACTTGCAGAGGAGGGCTCTGGTCTCTGAGGATGAGAGGTAGCTGA 1909
```





```
Qy        150 ---------CCTCCTCCAGCCGCGCCACCTGGTTCTGCAGGG-----------TCAGCA 112
                       |||   |||      |||||||||      |||           |||
Db       1145 AsnGlyLeuProGlyProIleGlyProProGlyProArgGlyArgThrGlyAspSerGly 1164

Qy        111 GCTGCTGCGCCCCCCTCATCCTCCTTTCGCCCGCCGGCCCCCTGCGCCGCCTGGTCCT 55
                |||   ||||||   ||||||          ||||||   ||||||||||||
Db       1165 ProAlaGlyProProGlyProPro---------GlyProProGlyProProGlyPro 1180
```

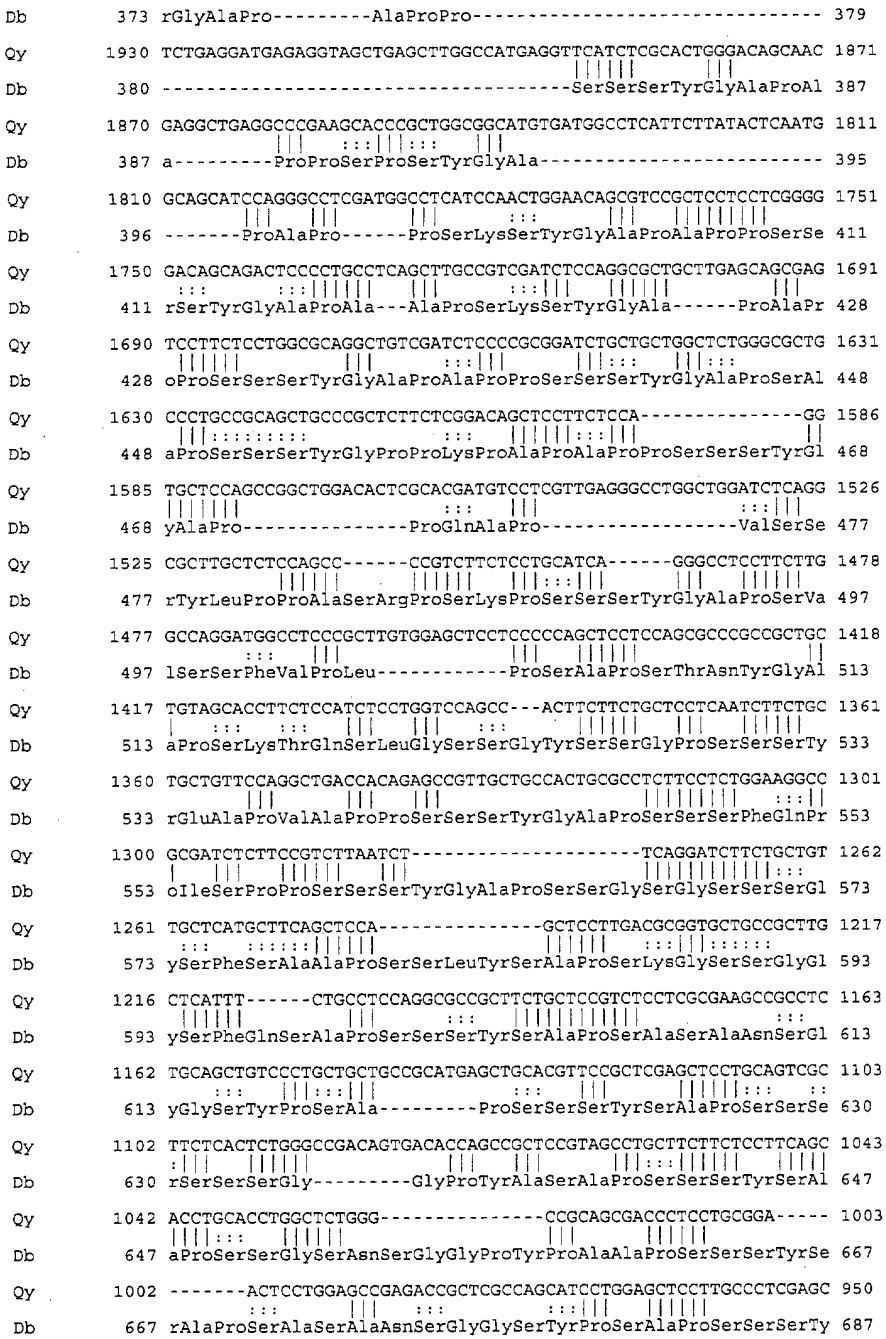
RESULT 8
Q9VE45
ID   Q9VE45   PRELIMINARY;   PRT;   950 AA.
AC   Q9VE45;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CG7709 protein.
GN   CG7709.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003723; AAF55584.1; -.
DR   FlyBase; FBgn0038642; CG7709.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   950 AA; 91280 MW; B57AB0EA4D30D3D7 CRC64;

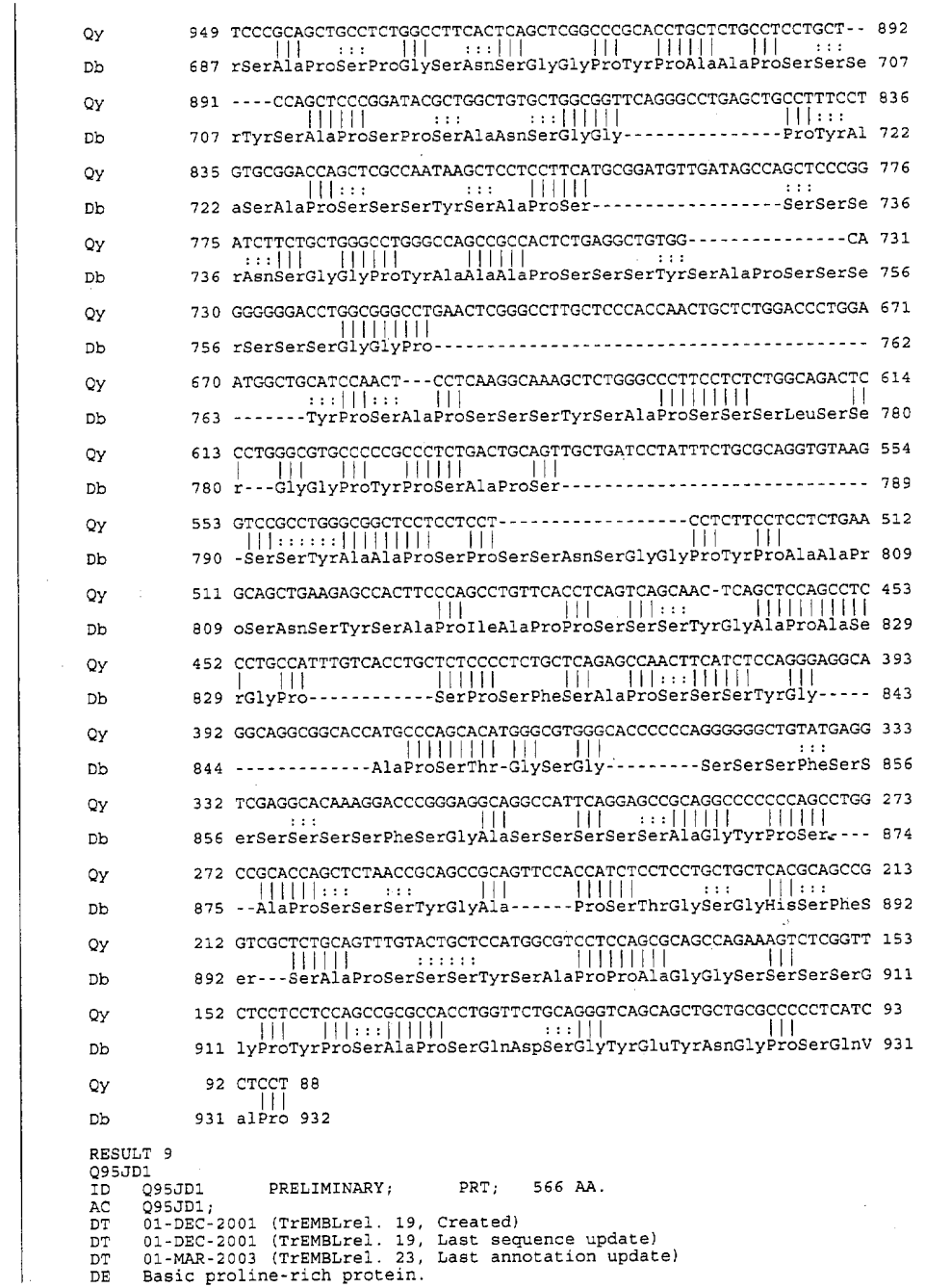
Alignment Scores:
Pred. No.: 4.79e-18 Length: 950
Score: 553.00 Matches: 282

Percent Similarity: 37.99% Conservative: 99
Best Local Similarity: 28.12% Mismatches: 380
Query Match: 9.68% Indels: 243
DB: 5 Gaps: 49

US-10-032-996-6 (1-3121) x Q9VE45 (1-950)

```
Qy       2854 ATGAGGGTCCAGTTCTTTTGG------GCCATGGCCCAAATTTGTTGATCCCAGTGAGGG 2801
              ||||||||||||   |||         |||:::|||   :::     |||
Db         93 MetArgValGlnLeuPheGlyValIleAlaValAlaValAlaValSerLeu------ 110

Qy       2800 TACAGATGAGGGCCTGGATTTAGGGTGTGCGGTAAGGACTGCCCTTCACAGAAGCAAAAC 2741
                                  |||      |||   :::   |||           |||
Db        111 -----------------ValArgAspValGlyAlaGluProPro------ValAsnAsn 122

Qy       2740 AGGCAGCTGCCCCTTTCAGCAGGCTCGGAGTCTCCCTCCAAGGCAGGGTCT--------- 2690
              ||||||   |||:::                |||||   |||:::
Db        123 AlaTyrLeuProProSerSerProGlnArgProSerSerLysTyrGlyAlaProProVal 142

Qy       2689 ---------GCCCCGAGGGCTTACAGGGGGTTTTTCCGGACATCAATCATCCCCGGGCTG 2639
                      |||   ||| :::|||                        |||:::|||
Db        143 SerSerTyrLeuProProAla-SerGlyProAlaProSerPheAsnSerAlaProSerSe 162

Qy       2638 GCTCGTCGCAGTTCCCGCCGGGGCTTGGACAAAGGCCCAAAGTTCCAGGGCAGGCCTGCC 2579
                 |||   |||                :::||||||:::   :::|||   ||
Db        162 rSerTyrAlaAlaPro-----------------SerGlnSerAlaSerSerGlyGlyPr 176

Qy       2578 TCACCCACAGGAAGCACCCGCCCCACCAGGGGCTCAGCCGCCTCCCGCTGCC-------- 2527
              |:::|||   ||||||   |||     :::     |||||||||:::
Db        176 oTyrProAlaAlaAlaProArgProSerSerSerTyrGlyProProAlaSerArgProSe 196

Qy       2526 -------------------------------TCAGTTCCTCGGGGGACCCCT-------- 2506
                                             |||       |||   |||
Db        196 rSerSerTyrGlyProProProSerArgProSerGlnSerTyrGlyProProProGlnAl 216

Qy       2505 ----GCTCCTCACCAC------ACAGGCTCGAGCGTTTCCAGGTCAAGGGTAACGGAGCG 2456
                  |||||||||        ||||||          |||
Db        216 aLysLysProHisHisArgArgProSerSerSerTyrGlyAlaProArgProAlaProPr 236

Qy       2455 TGGACCAAGTCCCGCGTCTCCTCCCGGGTGCGGG---GGGCCCCCTCAGTGAGGGGGGAC 2399
                  |||      :::||||||                |||||||||:::
Db        236 oSerGlnSerTyrGlyAlaProProSerAlaSerTyrGlyProProLysSerAlaProPr 256

Qy       2398 AGCCAGAGAAGCTCGGGTGCCAGGTGGAGCTCATCTTCATTTCCAGGAGCCTGTCTGCCC 2339
              ::::::          |||                    :::     |||
Db        256 oSerGlnSerTyrGlyAlaProAlaProProSerSerLysTyrGlyProProLysSerAl 276

Qy       2338 TCCGAGCACAGGCTCCTCTTCTCCCCACCCCTGCTGTGGCCTACAGCGTTCACACCGCCG 2279
              ||||||:::      :::|||  |||
Db        276 aProSerSerSerTyrGlyAlaProArgPro---------------------------- 286

Qy       2278 AGCTTCTGTTTCAGTTCCTGGTTTATCCACATGTAACGGCCCAGTTCCTTCTCCAGAGCT 2219
                    :::   |||     |||:::      |||   |||
Db        287 ----------AlaAlaProSer---SerSerTyrGlyAlaProAlaProProSerSerSe 302

Qy       2218 TGAATCCGGGCCTCATACTGCCTCCTGCTGTCTGCTAACCCTTCACCGAGGTGGTCTCGA 2159
                    |||     |||
Db        302 rTyrGlyAlaProAlaAlaProSerSerSerTyr------------------------- 313

Qy       2158 CTCTGCTGCAGGAGCAGCTGCATGTTCTGCTCGTGCTCCTTCTGCTGCAGGGTCAGCTGG 2099
              |||      |||  :::   |||       ||||||:::|||      |||:::
Db        314 -GlyAlaProAlaAlaProSerSerSerTyrGlyAlaProAlaAlaProSerSerSerTy 333

Qy       2098 CGGT-----------CCATCTCCAGGCGCTGCCGCTCCAGGGCCACCTCCAGCCAGTAC 2051
                |||           ||||||          ||||||   ||||||   :::|||
Db        333 rGlyAlaProAlaProProSerLysSerTyrGlyAlaProAlaProProSerSerSerTy 353

Qy       2050 ACCAGCCTCTGCTGCTCCTCCAGCTGCATCTCCAGTTCCGAGAAGGCAATCTGCTGCTGG 1991
              |||   |||||||||      :::          |||          |||::::::
Db        353 rGlyAlaProAlaAlaProSerLysSerTyrGlyAlaProAlaProProSerSerSerTy 373

Qy       1990 TGCTGCTCCTCTCGGAGCGTCACCACCTTGTCAAAATACTTGCAGAGGAGGGCTCTGGTC 1931
              ||||||         :::||||||
```





```
RESULT 9
Q95JD1
ID   Q95JD1      PRELIMINARY;      PRT;   566 AA.
AC   Q95JD1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Basic proline-rich protein.
```

```
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Parotid gland;
RA   Zhang Q., Szalay A.A., Kyeyune-Nyombi E., Sands J.F., Oberg K.C.,
RA   Tieche J.-M., Leonora J.;
RT   "Cloning and expression of a novel proline-rich protein from porcine
RT   parotid glands.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY035847; AAK61381.1; -.
DR   InterPro; IPR003882; Pistil_extensin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE   566 AA;  53213 MW;  E33B3B5E1BDEB81A CRC64;
```

Alignment Scores:

Pred. No.:	7.59e-18	Length:	566
Score:	548.50	Matches:	226
Percent Similarity:	26.82%	Conservative:	14
Best Local Similarity:	25.25%	Mismatches:	270
Query Match:	9.60%	Indels:	385
DB:	6	Gaps:	36

US-10-032-996-6 (1-3121) x Q95JD1 (1-566)

```
Qy       2706 CCTCCAAGGCAGGGTCTGCCCCGAGGGCTTACAGGGGGTTTTTCCGGACATCAATCATCC 2647
Db         43 ProProArgProProProProGluGluSerGlnGly------GluGlyHisGlnLysArg 60

Qy       2646 CCGGGCTGGCTCGTCGCAGTTCCCGCCGGGGCTTGGACAAAGGCCCAAAGTTCCAGGGCA 2587
Db         61 ProArg--------------ProProGly------------------------------ 65

Qy       2586 GGCCTGCCTCACCCACAGGAAGCACCCGCCCCACCAGGGGCTCAGCCGCCTCCCGCTGCC 2527
Db         66 ------AspGlyProGluGlnGlyProAlaProProGlyAlaArgProProPro------ 81

Qy       2526 TCAGTTCCTCGGGGGACCCCTGCTCCTCACCACACAGGCTCGAGCGTTTCCAGGTCAAGG 2467
Db         81 ------------------------------------------------------------ 81

Qy       2466 GTAACGGAGCGTGGACCAAGTCCCGCGTCTCCTCCCGGGTGCGGGGGGCCCCCTCAGTGA 2407
Db         82 ------------GlyProProProProGlyProPro------------------------ 89

Qy       2406 GGGGGGACAGCCAGAGAAGCTCGGGTGCCAGGTGGAGCTCATCTTCATTTCCAGGAGCCT 2347
Db         89 ------------------------------------------------------------ 89

Qy       2346 GTCTGCCCTCCGAGCACAGGCTCCTCTTCTCCCCACCCCTGCTGTGGCCTACAGCGTTCA 2287
Db         90 ------ProPro------------------------------------------------ 91

Qy       2286 CACCGCCGAGCTTCTGTTTCAGTTCCTGGTTTATCCACATGTAACGGCCCAGTTCCTTCT 2227
Db         92 ----------------------------------------------GlyProAlaPro--- 95

Qy       2226 CCAGAGCTTGAATCCGGGCCTCATACTGCCTCCTGCTGTCTGCTAACCCTTCACCGAGGT 2167
Db         95 ------------------------------------------------------------ 95

Qy       2166 GGTCTCGACTCTGCTGCAGGAGCAGCTGCATGTTCTGCTCGTGCTCCTTCTGCTGCAGGG 2107
Db         96 ----------------------------------------------ProGlyAlaArgPro 100

Qy       2106 TCAGCTGGCGGTCCATCTCCAGGCGCTGCCGCTCCAGGGCCA---CCTCCAGCCAGTACA 2050
Db        101 ProProGlyProProProProGlyProProProProGlyProAlaProProGlyAlaArg 120
```

```
Qy       2049 CCAGCCTCTGCTGCTCCTCCAGCTGCATCTCCAGTTCCGAGAAGGCAATCTGCTGCTGGT 1990
Db        121 ProProProGlyProProProProGlyProProProPro----------------Gly 134

Qy       1989 GCTGCTCCTCTCGGAGCGTCACCACCTTGTCAAAATACTTGCAGAGGAGGGCTCTGGTCT 1930
Db        135 ProAlaProProGlyAlaArgProPro---------------------------------- 143

Qy       1929 CTGAGGATGAGAGGTAGCTGAGCTTGGCCATGAGGTTCATCTCGCACTGGGACAGCAACG 1870
Db        143 ------------------------------------------------------------ 143

Qy       1869 AGGCTGAGGCCCGAAGCACCCGCTGGCGGCATGTGATGGCCTCATTCTTATACTCAATGG 1810
Db        144 ---ProGlyProProProProAlaGlyGlyLeu---------GlnGlnGlyProAlaPro 159

Qy       1809 CAGCATCCAGGGCCTCGATGGCCTCATCCAACTGGAACAGCGTCCGCTCCTCCTCGGGGG 1750
Db        160 SerHisValGlyProLysLysLysProProProProGlyAlaGlyHisProProArg--- 178

Qy       1749 ACAGCAGACTCCCCTGCCTCAGCTTGCCGTCGATCTCCAGGCGCTGCTTGAGCAGCGAGT 1690
Db        179 ---------ProProProProAlaAsnGluSerGlnProGly----------------- 189

Qy       1689 CCTTCTCCTGGCGCAGGCTGTCGATCTCCCCGCGGATCTGCTGCTGGCTCTGGGCGCTGC 1630
Db        190 ProArgProProProGlyProProSerProProAlaAsnAspSerGlnGluGlySerPro 209

Qy       1629 CCTGCCGCAGCTGCCCGCTCTTCTCGGACAGCTCCTTCT--------------------- 1591
Db        210 ProSerAlaAspGlyProGlnGlnGlyProAlaProSerGlyAspLysProLysLysLys 229

Qy       1590 CCAGGTGCTCCAGCCGGCTGGACACTCGCACGATGTCCTCGTTGAGGGCCTGGCTGGATC 1531
Db        230 ProProProProAlaGly-----------------ProProProProProProProPro 243

Qy       1530 TCAGGCGCTTGCTCTCCAGCCCCGTCTTCTCCTGCATCAGGGCCTCCTTCTTGGCCAGGA 1471
Db        244 ProGlyProProProProGlyProAlaProProGlyAlaArgProPro------ProGly 261

Qy       1470 TGGCCTCCCGCTTGTGGAGCTCCTCCCCCAGCTCCTCCAGCGCCCGCCGCTGCTGTAGCA 1411
Db        262 ---ProProProProGlyProProProProGlyProAlaProProGlyAlaArgProPro 280

Qy       1410 CCTTCTCCATCTCCT------GGTCCAGCCACTTCTTCTGCTCCTCAATCTTCTGCTGCT 1357
Db        281 ProGlyProProProProGlyProProProProGlyProAlaProHisGlyAlaArgPro 300

Qy       1356 GTTCCAGGCTGACCACAGAGCCGTTGCTGCCACTGCGCCTCTTCCTCTGGAAGGCCGCGA 1297
Db        301 ProProGlyProPro-----------------------------------------ProPro 307

Qy       1296 TCTCTTCCGTCTTAATCTTCAGGATCTTCTGCTGTTGCTCATGCTTCAGCTCCAGCTCCT 1237
Db        308 GlyProPro-----------------------------------------ProProGlyPro 314

Qy       1236 TGACGCGGTGCTGCCGCTTGCTCATTTCTGCCTCCAGGCGCCGCTTCTGCTCCGTCTCCT 1177
Db        315 --------------------------AlaProProGlyAlaArgProProProGlyPro 325

Qy       1176 CGCGAAGCCGCCTCTGCAGCTGTCCCTGCTGCTGCCGCATGAGCTGCACGTTCCGCTCGA 1117
Db        325 ------------------------------------------------------------ 325

Qy       1116 GCTCCTGCAGTCGCTTCTCACTCTGGGCCGACAGTGACACCAGCCGCTCCGTAGCCTGCT 1057
Db        326 --------------------------ProProProGlyProProProProGlyProAla 336

Qy       1056 TCTTCTCCTTCAGCACCTGCACCTGGCTCTGGGCCGCAGCGACCCTCCTGCGGAACTCCT 997
Db        337 ProProGlyAlaArgProProProGlyProProProProGlyProProProProGlyPro 356

Qy        996 GGAGCCGAGACCGCTCGCCAGCATCCTGGAGCTCCTTGCCCTCGAGCTCCCGCAGCTGCC 937
```

```
                        ||||||      ||||||   |||   |||      |||
Db        357 AlaProProGlyAlaArgProProProGlyProProProGlyProProProProGly 376

Qy        936 TCTGGCCTTCACTCAGCTCGGCCCGCACCTGCTCTGCCTCCTGCTCCAGCTCCCGGATAC 877
                          |||||||||           ||||||   |||   |||
Db        377 ProAlaProProGlyAlaArgPro------------ProProGlyProProPro------ 390

Qy        876 GCTGGCTGTGCTGGCGGTTCAGGGCCTGAGCTGCCTTTCCTGTGCGGACCAGCTCGCCAA 817
                                  ||||||      |||         |||||||||
Db        391 ------------------ProGlyProProProPro---------GlyProAlaProPro 401

Qy        816 TAAGCTCCTCCTTCATGCGGATGTTGATAGCCAGCTCCCGGATCTTCTGCTGGGCCTGGG 757
              |||   |||      |||         |||   |||                  ||||||
Db        402 GlyAlaArgProProProGly---------ProProProProGlyProProProProGly 418

Qy        756 CCAGCCGCCACTCTGAGGCTGTGGCAGGGGGGACCTGGCGGG---CCTGAACTCGGGCCT 700
              ||||||                        ||||||      |||      ||||||
Db        419 ProAlaPro-----------------------ProGlyAlaArgProLeuProGlyPro 430

Qy        699 TGCTCCCACCAACTGCTCTGGACCCTGGAATGGCTGCATCCAACTCCTCAAGGCAAAGCT 640
                                                  |||   |||
Db        431 ------------------------------------ProProPro------------ 433

Qy        639 CTGGGCCCTTCCTCTCTGGCAGACTCCCTGGGCGTGCCCCCGCCCTCTGACTGCAGTTGC 580
                                        |||   |||||||||
Db        434 ----------------------------GlyProProProPro--------------- 438

Qy        579 TGATCCTATTTCTGCGCAGGTGTAAGGTCCGCCTGGGCGGCTCCTCCTCCTCCTCTTCCT 520
                          |||                 |||   ||||||||||||   |||
Db        439 ------------------GlyProAlaProProGlyAlaArgProProProProProPro 452

Qy        519 CCTCTGAAGCAGCTGAAGAGCCACTTCCCAGCCTGTTCACCTCAGTCAGCAACTCAGCTC 460
              |||                       ||||||   |||              :::
Db        453 ProProAlaAspGluProGlnGlnGlyProAlaProSerGlyAspLysProLysLysLys 472

Qy        459 CAGCCTCCCTGCCATTTGTCACCTGCTCTCCCCTCTGCTCAGAGCCAACTTCATCTCCAG 400
                ||||||          ||||||   |||
Db        473 ProProPro------------ProAlaGlyPro-------------------------- 479

Qy        399 GGAGGCAGGCAGGCGGCACCATGCCCAGCACATGGGCGTGGGCACCCCCCAGGGGGGCTG 340
                              |||   |||          |||   ||||||   |||
Db        480 ---------------ProProProProProProProProGlyProProProProGly--- 493

Qy        339 TATGAGGTCGAGGCACAAAGGACCCGGGAGGCAGGCCATTCAGGAGCCGCAGGCCCCCCC 280
                                                  |||                 |||||||||
Db        494 ---------------------ProAlaProProGlyAlaArgProArgProGlyProPro 506

Qy        279 AGCCTGGCCGCACCAGCTCTAACCGCAGCCGCAGTTCCACCATCTCCTCCTGCTGCTCAC 220
                                                 ||||||   ||||||
Db        507 ----------------------------------ProProGlyProProProProGly 514

Qy        219 GCAGCCGGTCGCTCTGCAGTTTGTACTGCTCCATGGCGTCCTCCAGCGCAGCCAGAAAGT 160
              |||          |||                 |||||||||       |||
Db        515 ProAlaProProGlyAla-----------------ArgProProProGlyProProPro 528

Qy        159 CTCGGTTCTCCTCCTCCAGCCGCGCCACCTGGTTCTGCAGGGTCAGCAGCTGCTGCGCCC 100
              |||   |||||||||         |||       |||                  |||
Db        529 ProGlyProProProProGlyProAlaProProGlyAlaArgProProProGlyProPro 548

Qy         99 CCTCATCCTCCTTTCGCCCGCCGGCCCCCTGCGCCGCCTGGTCCT 55
              |||   ||||||          |||      ||||||||||||
Db        549 ProProProProGlyProSerProProArgProProProGlyPro 563

RESULT 10
Q8IR84
ID   Q8IR84      PRELIMINARY;      PRT;  1099 AA.
AC   Q8IR84;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG32656-PA.
GN   CG32656.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003488; AAF48140.3; -.
SQ   SEQUENCE   1099 AA;  94352 MW;  1A24A01379BA0DA1 CRC64;

Alignment Scores:
Pred. No.:              1.64e-17      Length:         1099
Score:                  541.50        Matches:        299
Percent Similarity:     35.15%        Conservative:   69
Best Local Similarity:  28.56%        Mismatches:     452
Query Match:            9.48%         Indels:         227
DB:                     5             Gaps:           50

US-10-032-996-6 (1-3121) x Q8IR84 (1-1099)

Qy       2832 CCATGGCCCAAATTTGTTGATCCC-----------AGTGAGGGTACAGATGAGGGCCTG 2785
Db          6 ProLeuProGlyAlaProGluProPheCysProGlySerArgGlyLeuGlnValGlyGly 25

Qy       2784 GATTTAGGGTGTGCGGTAAGGACTGCCCTTCACAGAAGCAAAACAGGCAGCTGCCCCTTT 2725
Db         26 ThrIleGlyMetPheGlyArg-------------------------GlyCysGlnGly 36

Qy       2724 CAGCAGGCTCGGAGTCTCCCTCCAAGGCAGGGTCTGCCCCGAGGGCTTACA--------- 2674
Db         37 GlnGlu------ProAlaProProGlySerThrGlyIleAspGlyThrThrGluProAla 54

Qy       2673 GGGGGTTTTTCCGGACATCAATCATCCCCGGGCTGGCTCGTCGCA---------GTTCCC 2623
Db         55 GlyGlyPheGlyProArgThrAspProGlyGluLeuProAlaGlyThrLeuGluPro 74

Qy       2622 GCCGGGGCTTGGACAAAGGCCCAAAGTTCCAGGGCAGGCCTGCCTCACCCA--------- 2572
Db         75 AlaGlyAlaProProAlaGlyThrLeuGluSerAlaGlyAlaProProAlaGlyThrLeu 94

Qy       2571 CAGGAAGCACCCGCCCCACCAGGGGCTCAGCCGCCTCCCGCTGCCTCAGTTCCTCGGGGG 2512
Db         95 GluSerAlaGlyAlaProProAlaGlyThrLeuGluProAlaGlyAlaProProAlaGly 114

Qy       2511 ACCCCTGCTCCTCACCACACAGGCTCGAGCGTTTCCAGGTCAAGGGTAACGGAGCGTGGA 2452
Db        115 Thr------LeuGluSerAlaGlyAlaProProAlaGlyThrLeuAlaSerAlaGlyAla 132

Qy       2451 CCAAGTCCCGCGTCTCCTCCCGGGTGCGGGGGGCCCCCTCAGTGAGGGGGGACAGCCAGA 2392
Db        133 ProProProGlyThrLeuGluSerAlaGlyAlaProPro------AlaGlyThrLeuAla 150

Qy       2391 GAAGCTCGGGTGCCAGGTGGAGCTCATCTTCATTTCCAGGAGCCTGTCTGCCCTCCGAGC 2332
Db        151 SerAlaGlyAlaPro-----------------------------------ProAlaGly 158

Qy       2331 ACAGGCTCCTCTTCTCCCCACCCCTGCTGTGGCCTACAGCGTTCACACCGCCGAGCTTCT 2272
Db        159 ThrLeuGluSerAlaGlyAlaProProAlaGlyThrLeuGluSer---------AlaGly 175

Qy       2271 GTTTCAGTTCCTGGTTTATCCACATGTAACGGCCCAGTTCCTTCTCCAGAGCTTGAATCC 2212
Db        176 AlaProProAlaGlyThrLeuGluSerAlaGlyAlaProProAlaGlyThrLeuGluSer 195

Qy       2211 ---GGGCCTCATACTGCCTCCTGCTGTCTGCTAACCCTTCACCGAGGTGGTCTCGACTCT 2155
Db        196 AlaGlySerProProAlaGly-----------ThrLeuGluProAlaGly---AspPro 210

Qy       2154 GCTGCAGGAGCAGCTGCATGTTCTGCTCGTGCTCCTTCTGCTGCAGGGTCAGCTGGCGGT 2095
Db        211 ProAlaGlyGlyAspPro---SerAlaGlyGluProGlyAlaAlaGluAspProSerAla 229

Qy       2094 CCATCTCCAGGCGCTGCC-----------GCTCCAGGGCCACCTCCAGCCAGTACACCA 2047
Db        230 GlyAlaProGlyAlaAlaValValThrSerAlaGlyGluProGlyAlaAlaGluAspPro 249

Qy       2046 GCCTCTGCTGCTCCTCCAGCTGCATCTCCAGTTCCGAGAAGGCAATCTGCTGCTGGTGCT 1987
Db        250 SerAlaGlyAlaProGlyAlaAlaValValThrSerAlaGlyGluProGlyAlaAlaGlu 269

Qy       1986 GCTCCTCTCGGAGCGTCACCACCTTGTCAAAATACTTGCAGAGGAGGGCTCTGGTCTCTG 1927
Db        270 AspProSerAlaGlyAlaProGlyAlaAlaAspAspProSerAlaGlyAla--------- 286

Qy       1926 AGGATGAGAGGTAGCTGAGCTTGGCCATGAGGTTCATCTCGCACTGGGACAGCAACGAGG 1867
Db        287 ----------------------ProGlyAlaAlaValValThrSerAlaGlyAlaPro 298

Qy       1866 CTGAGGCCCGAAGCACCCGCTGGCGGCATGTGATGGCCTCATTCTTATACTCAATGGCAG 1807
Db        299 GlyAlaAlaGluAspProSerAlaGly------------------------------- 307

Qy       1806 CATCCAGGGCCTCGATGGCCTCATCCAACTGGA----------------------ACA 1771
Db        308 GluProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValValThrSer 327

Qy       1770 GCGTCCGCTCCTCCTCGGGGGACAGCAGACTCCCCTGCCTCAGCTTGCCGTCGATCTCCA 1711
Db        328 AlaGlyAlaPro------GlyAlaAlaAspAspProSerAlaGly---------AlaPro 342

Qy       1710 GGCGCTGCTTGAGCAGCGAGTCCTTCTCCTGGCGCAGGCTGTCGATCTCCCCGCGGATCT 1651
Db        343 GlyAlaAlaValValThrSerAlaGlyGluProGlyAlaAlaGluAspProSerAlaGly 362

Qy       1650 GCTGCTGGCTCTGGGCGCTGCCCTGCCGCAGCTGCCCGCTCTTCTCGGACAGCTCCTTCT 1591
Db        363 AlaProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValValThrSer 382

Qy       1590 CCAGGTGCTCCAGCCGGCTGGACACTCGCACGATGTCCTCGTTGAGGG---CCTGGCTGG 1534
Db        383 AlaGlyAlaProGlyAla---------AlaGluAspProSerAlaGlyGluProGly--- 398

Qy       1533 ATCTCAGGCGCTTGCTCTCCAGCCCCGTCTTCTCCTGCATCAGGGCCTCCTTCTTGGCCA 1474
Db        399 ------AlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValValThrSerAla 416

Qy       1473 GGATGGCCTCCCGCTTGTGGAGCTCCTCCCCCAGCTCCTCCAGCGCCCGCCGCTGCTGTA 1414
Db        417 GlyGluProGlyAla------AlaGluAspProSerAlaGlyAlaProGlyAlaAlaVal 434

Qy       1413 --------------------------GCACCTTCTCCATCTCCTGGTCCAGCCACTTCT 1381
Db        435 ValThrSerAlaGlyGluProGlyAlaAlaGluAspProSerAlaGly---AlaProGly 453

Qy       1380 TCTGCTCCTCAATCTTCTGCTGCTGTTCCAGGCTGACCACAGAGCCGTTGCTGCCACTGC 1321
Db        454 AlaAlaAspAspProSerAlaGlyAlaProGly-----------------AlaAlaVal 467

Qy       1320 GCCTCTTCCTCTGGAAGGCCGCGATCTCTTCCGTCTTAATCTTCAGGATCTTCTGCTGTT 1261
Db        468 ValThrSerAlaGlyGluProGlyAlaAlaGluAspProSerAlaGlyAlaProGlyAla 487

Qy       1260 GCT---CATGCTTCAGCTCCAGCTCCTTGACGCGGTGCTGCC---GCTTGCTCATTTCTG 1207
Db        488 AlaAspAspProSerAlaGlyAlaPro------GlyAlaAlaValValThrSerAlaGly 505

Qy       1206 CCTCCAGGCGCCGCTTCTGCTCCGTCTCCTCGCGAAGCCGCCTCTGCAGCTGTCCCTGCT 1147
Db        506 GluProGlyAlaAlaGluAspProSerAlaGlyGluProGlyAlaAlaAspAspProSer 525

Qy       1146 GCTGCCGCATGAGCTGCACGT-------------------------TCCGCTCGAGCT 1114
Db        526 AlaGlyAlaProGlyAlaAlaValValThrSerAlaGlyGluProGlyAlaAlaGluAsp 545

Qy       1113 CCTGCAGTCGCTTCTCACTCTGGGCCGACAGTGACACCAGCCGCT---CCGTAGCCTGCT 1057
```

```
Db        546 ProSerAlaGlyAlaProGlyAlaAlaValValThrSerAlaGlyGluProGlyAlaAla 565

Qy       1056 TCTTCTCCTTCAGCACCTGCACCTGGCTCTGGGCCGCAGCGACCCTCCTGCGGAACTCCT 997

Db        566 GluAspProSerAlaGlyAlaProGlyAlaAlaAspAsp---ProSerAlaGlyAlaPro 584

Qy        996 GGAGCCGAG------ACCGCTCGCCAGCATCCTGGAGCT---------- 964

Db        585 GlyAlaAlaValValThrSerAlaGlyGluProGlyAlaAlaGluAspProSerAlaGly 604

Qy        963 ----------CCTTGCCCTCGAGCTCCCGCAGCTGCCTCTGGCCTTCAC 925

Db        605 GluProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValValThrSer 624

Qy        924 TCAGCTCGGCCCGCACCTGCTCTG---CCTCCTGCTCCAGCTCCCGGATACGCTGGC--- 871

Db        625 AlaGlyGluProGlyAlaAlaGluAspProSerAlaGlyAlaProGlyAlaAlaAspAsp 644

Qy        870 ---TGTGCTGGCGGTTCAGGGCCTGAGCTGCCTTTCCTGTGCGGACCAGCTCGCCAATAA 814

Db        645 ProSerAlaGlyAlaProGlyAlaAlaValValThrSerAlaGlyGluProGlyAlaAla 664

Qy        813 GCTCCTCCTTCATGCGGATGTTGATAGCCAGCTCCCGGATCTTCTGCTGGGCCT------ 760

Db        665 GluAspProSerAlaGly-----------AlaProGlyAlaAlaAspAspProSerAla 680

Qy        759 ---GGGCCAGCCGCCACTCTGAGGCTGTGGCAGGGGGGACCTGGCGGGCCTGAACTCGGG 703

Db        681 GlyAlaProGlyAlaAlaValValThrSerAlaGlyAlaProGlyAlaAlaGlu------ 698

Qy        702 CCTTGCTCCCACCAACTGCTCTGGACCCTGGAATGGCTGCATCCAACTCCTCAAGGCAAA 643

Db        699 ------Asp--ProSerAlaGlyGluProGlyAlaAlaAspAspProSerAla------- 713

Qy        642 GCTCTGGGCCCTTCCTCTCTGGCAGACTCCCTGGGCGTGCCCCCGCCCTCTGACTGCAGT 583

Db        714 ----------------------GlyAlaProGlyAlaAlaValVal------------- 721

Qy        582 TGCTGATCCTATTTCTGCGCAGGTGTAAGGTCCGCCTGGGCGGCTCCTCCTCCTCCTCTT 523

Db        722 ------------------------------------------------ThrSerAlaG 725

Qy        522 CCTCCTCTGAAGCAGCTGAAGAGCCA---------CTTCCCAGCCTGTTCACCTCAGTCA 472

Db        725 lyGluProGlyAlaAlaGluAspProSerAlaGlyAlaProGlyAlaAlaAspAspProS 745

Qy        471 GCAACTCAGCTCCAGCCTCCCTGCCATTTGTCACCTGCTCTCCCCTCTGCTCAGAGCCAA 412

Db        745 erAlaGlyAlaProGlyAlaAlaValValThrSerAlaGlyGluProGlyAlaAlaGluA 765

Qy        411 CTTCATCTCCAGGGAGGCAGGCAGGCGGCACCATGCCCAGCACATGGGCGTGGGCACCCC 352

Db        765 spSerSerAlaGlyGluProGlyAlaAlaAspAspProSer------AlaGlyAlaProG 783

Qy        351 CCAGGGGGGCTGTATGAGGTCGAGGC----------ACAAAGGACCCGGGAGGCAGGC 304

Db        783 lyAlaAlaValValThrSerAlaGlyGluProGlyAlaAlaGluAspPro-SerAlaGly 802

Qy        303 CATTCAGGAGCCGCAGGCCCCCCCAGCCTGGCCGCACCA---------GCTCTAACCGCA 253

Db        803 AlaProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValValThrSer 822

Qy        252 GCCGCAGTTCCA----------CCATCT----------CCT 232

Db        823 AlaGlyGluProGlyAlaAlaGluAspProSerAlaGlyAlaProGlyAlaAlaAspAsp 842

Qy        231 CCTGCTGCTCACGCAGCCGGTCGCTCTGCAGTTTGTACTGCTCCATGGCGTCCTCCAGCG 172

Db        843 ProSerAlaGlyAlaProGlyAlaAlaValValThrSerAlaGlyAlaProGlyAlaAla 862

Qy        171 CAGCCAGAAAGTCTCGGTTCTCCT----------CCTCCAGCCGCGCCACCTGGT 127

Db        863 GluAspProSerAlaGlyGluProGlyAlaAlaAspAspProSerAlaGlyAlaProGly 882

Qy        126 TCTGCAGGGTCAGCAGCTGCTGCGCCCCCT---------CATCCTCCTTTCGCCCGCCGG 76

Db        883 AlaAlaValValThrSerAlaGlyGluProGlyAlaAlaGluAspProSerAlaGlyAla 902

Qy         75 CCCCCTGCGCCGCCTGGTCCTCGACGGAGGCGCTCTCGATGCCGCTATCGGGCCCGGAGG 16

Db        903 ProGlyAlaAlaAspAspPro-SerAlaGlyAlaProGlyAlaAlaValValThrSerAl 922

Qy         15 CGGAGCTCAGGGC 3

Db        922 aGlyAlaProGly 926
```

RESULT 11
O76045
ID   O76045   PRELIMINARY;   PRT; 1461 AA.
AC   O76045;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Pro alpha 1(I) collagen.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85130970; PubMed=2857713;
RA   Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT   Promoter structure, AluI repeats, and polymorphic transcripts.";
RL   J. Biol. Chem. 260:2315-2320(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88329734; PubMed=2843432;
RA   D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT   "Complete nucleotide sequence of the region encompassing the first
RT   twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL   Gene 67:105-115(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89025644; PubMed=3178743;
RA   Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA   Jaenisch R., Prockop D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT   of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91138770; PubMed=1995349;
RA   Maatta A., Bornstein P., Penttinen R.P.;
RT   "Highly conserved sequences in the 3'-untranslated region of the
RT   COL1A1 gene bind cell-specific nuclear proteins.";
RL   FEBS Lett. 279:9-13(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92157916; PubMed=1787829;
RA   Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA   Olsen A.S., Prockop D.J.;
RT   "Completion of the last half of the structure of the human gene for
RT   the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL   Matrix 11:375-379(1991).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98107942; PubMed=9443882;
RA   Korkko J., Ala-Kokko L., De Paepe A., Nuytinck L., Earley J.,
RA   Prockop D.J.;
RT   "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT   scanning by conformation-sensitive gel electrophoresis identifies only
RT   COL1A1 mutations in 15 patients with osteogenesis imperfecta type I:
RT   identification of common sequences of null-allele mutations.";

```
RL   Am. J. Hum. Genet. 62:98-110(1998).
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Korkko J.M., Earley J.J., Nuytinck L., DePaepe A., Prockop D.J.,
RA   Ala-Kokko L.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF017178; AAB94054.2; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Collagen; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Collagen.
SQ   SEQUENCE   1461 AA;  138630 MW;  9ACF6DE30EA78E21 CRC64;

Alignment Scores:
Pred. No.:                1.84e-17      Length:          1461
Score:                    540.50        Matches:         290
Percent Similarity:       30.49%        Conservative:    57
Best Local Similarity:    25.48%        Mismatches:      424
Query Match:              9.46%         Indels:          367
DB:                       4             Gaps:            58

US-10-032-996-6 (1-3121) x O76045 (1-1461)

Qy       2847 TCCAGTTCTTTTGGGCCATGGCCCAAATTTGTTGATCCCAGTGAGGGTACAGATGAGGGC 2788
Db        211 SerGlyProMetGlyProArgGlyProProGlyProGlyLysAsnGlyAspAspGly 230

Qy       2787 CTGGATTTAGGGTGTGCGGTAAGGACTGCCCTTCACAGAAGCAAAACAGGCAGCTGCCCC 2728
Db        231 ---GluAlaGlyLysProGlyArgProGly---GluArgGlyProProGlyPro------ 246

Qy       2727 TTTCAGCAGGCTCGGAGTCTCCCTCCAAGGCAGGGTCTGCCCCGAGGGCTTACAGGG--- 2671
Db        247 ---GlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeuPro---GlyMetLysGlyHis 264

Qy       2670 ---GGTTTTTCCGGACATCAATCATCCCCGGGCTGGCTCGTCGCAGTTCCCGCCGGGGCT 2614
Db        265 ArgGlyPheSerGlyLeuAspGlyAlaLysGly------AspAlaGlyProAlaGly--- 281

Qy       2613 TGGACAAAGGCCCAAAGTTCCAGGGCAGGCCTGCCTCACCCACAGGAAGCACCCGCCCCA 2554
Db        282 ----------------------------------------ProLysGlyGluProGlySer 288

Qy       2553 CCAGGGGCTCAGCCGCCTCCCGCTGCCTCAGTTCCTCGGGGGACCCCT------------ 2506
Db        289 ProGlyGluAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyGluArgGly 308

Qy       2505 ---------GCTCCTCACCACACAGGCTCGAGCGTTTCCAGGTCAAGGGTAACGGAGCGT 2455
Db        309 ArgProGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaAla 328

Qy       2454 GGACCA------------AGTCCCGCGTCTCCTCCCGGGTGCGGGGGGCCCCCTCAGTGA 2407
Db        329 GlyProProGlyProThrGlyProAlaGlyProProGlyPheProGlyAlaValGlyAla 348

Qy       2406 GGGGGGACAGCC--------------AGAGAAGCTCGGGTGCCAGGTGGAGCTCATCTT 2362
Db        349 LysGlyGluAlaGlyProGlnGlyProArgGlySerGluGlyProGlnGlyValArg--- 367

Qy       2361 CATTTCCAGGAGCCTGTCTGCCCTCCGAGCACAGGCTCCTCTTCTCCC------CACCCC 2308
Db        368 ------GlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnPro 385

Qy       2307 TGCTGTGGCCTACAGCGTTCACACCGCCGAGCTTCT--------------GTTTCAGTT 2263
Db        386 GlyAlaAspGlyGlnProGlyAlaLysGlyAlaAsnGlyAlaProGlyIleAlaGlyAla 405

Qy       2262 CCTGGTTTATCCACATGTAACGGCCCA----------------------GTTCCTTCT 2227
Db        406 ProGlyPheProGlyAlaArgGlyProSerGlyProGlnGlyProGlyGlyProProGly 425

Qy       2226 CCAGAGCTTGAATCCGGGCCTCATACTGCCTCCTGCTGTCTGCTAACCCTTCACCGAGGT 2167
Db        426 ProLysGlyAsnSerGlyGluProGlyAlaPro---------------------------- 436

Qy       2166 GGTCTCGACTCTGCTGCAGGAGCAGCTGCATGTTCTGCTCGTGCTCCTTCTGCTGCAGGG 2107
Db        437 GlySerLysGlyAspThrGlyAlaLysGly------------GluProGlyProValGly 452

Qy       2106 TCAGCTGGCGGTCCATCTCCAGGC---------------------------------- 2083
Db        453 ValGlnGlyProProGlyProAlaGlyGluGluGlyLysArgGlyAlaArgGlyGluPro 472

Qy       2082 ------GCTGCCGCTCCAGGGCCACCT---CCAGCCAGTACACCAGCCTCTGCTGCTCCT 2032
Db        473 GlyProThrGlyLeuProGlyProProGlyGluArgGlyGlyProGlySerArgGlyPhe 492

Qy       2031 CCAGCTGCA--------------------TCTCCAGTTCCGAGAAGGCAATCTGCTGCT 1993
Db        493 ProGlyAlaAspGlyValAlaGlyProLysGlyProAlaGlyGluArgGlySerProGly 512

Qy       1992 GGTGCTGCTCCTCTCGGAGCGTCACCACCTTGTCAAAATACTTGCAGAGGAGGGCTCTGG 1933
Db        513 ProAlaGlyProLysGlySerProGlyGluAlaGlyArgProGlyGluAlaGlyLeuPro 532

Qy       1932 TCTCTGAGG---ATGAGAGGTAGCTGAGCTTGGCCATGAGGTTCATCTCGCACTGGGACA 1876
Db        533 GlyAlaLysGlyLeuThrGlySerProGlySerProGlyProAspGlyLysThrGly--- 551

Qy       1875 GCAACGAGGCTGAGGCCCGAAGCACCCGCTGGCGGCATGTGATGGCCTCATTCTTATACT 1816
Db        552 ---------------ProProGlyProAlaGly---------------------------- 557

Qy       1815 CAATGGCAGCATCCAGGGCCTCGATGGCCTCATCCAACTGGAACAGCGTCCGCTCCT--- 1759
Db        558 GlnAspGlyArgProGlyPro------ProGlyProProGlyAlaArgGlyGlnAlaGly 575

Qy       1758 -----------------CCTCGGGGGACAGCAGACTCCCCTGCCTCAGCTTGCCGTCGA 1717
Db        576 ValMetGlyPheProGlyProLysGlyAlaAlaGlyGluProGlyLysAlaGlyGluArg 595

Qy       1716 TCT------------CCAGGCGCTGCTTGAGCAGCGAGTCCTTCTCCTGGCGCAGGCTGT 1669
Db        596 GlyValProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGluAlaGlyAla 615

Qy       1668 CGATCTCCCCGCGGATCTGCTGCTGGCTCTGGGCGCTGCCCTGCCGCAGCTGCCCGCTCT 1609
Db        616 GlnGlyPro---------------------------------------------------- 618

Qy       1608 TCTCGGACAGCTCCTTCTCCAGGTGCTCCAGCCGGCTGGACACTCGCACGATGTCCTCGT 1549
Db        619 ------------ProGlyProAlaGlyProAlaGlyGluArgGlyGluGlnGlyProAla 634

Qy       1548 TGAGGGCCTGGCTGGATCTCAGGCGCTTGCTCTCCAGCCCCGTCTTCTCCTGCATCAGGG 1489
Db        635 GlySerProGlyPhe---GlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGly 653

Qy       1488 CCTCCTTCTTGGCCAGGATGGCCTCCCGCTTGTGGAGCTCCTCCCCCAGCT--------- 1438
Db        654 LysProGlyGluGlnGlyValProGlyAspLeuGlyAlaProGlyProSerGlyAlaArg 673

Qy       1437 ------------------------------------CCTCCAGCGCCCGCCGCTGCT 1417
Db        674 GlyGluArgGlyPheProGlyGluArgGlyValGlnGlyProProGlyProAlaGlyPro 693

Qy       1416 GTAGCACCTTCTCCATCTCCTGGTCCAGCCACTTCTTCTGCTCCTCAATCTTCTGCTGCT 1357
Db        694 ArgGlyAlaAsnGlyAlaProGly---------AsnAspGlyAlaLysGlyAspAlaGly 710
```

```
Qy       1356 GTTCCAGGCTGACCACAGAGCCGTTGCTGCCACTGCGCCTCTTCCTCTGGAAGGCCG--- 1300
Db        711 AlaProGlyAlaPro----------------------GlySerGlnGlyAlaProGly 722

Qy       1299 ---CGATCTCTTCCGTCTTAATCTTCAGGATCTTCTGCTGTTGCTCATGCTTCAGCTCCA 1243
Db        723 LeuGlnGlyMetPro-----------GlyGluArgGlyAlaAlaGlyLeuProGlyPro 738

Qy       1242 GCTCCTTGACGCGGTGCTGCCGCTTGCTCA---------------------------- 1213
Db        739 LysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGly 758

Qy       1212 --------------------TTTCTGCCTCCAGGCGCCGCTTCTGCTCCGTCTCCTCGC 1174
Db        759 ValArgGlyLeuThrGlyProIleGlyProProGlyProAlaGlyAlaProGlyAspLys 778

Qy       1173 GAAGCCGCCTCTGCAGCTGTCCCTGCTGCTGCCGCATGAGCTGCACGTTCCGCTCGAGCT 1114
Db        779 GlyGluSerGlyProSerGlyProAla------------------------------- 787

Qy       1113 CCTGCAGTCGCTTCTCACTCTGGGCCGACAGTGACACCAGCCGCTCCGTAGCCTGCTTCT 1054
Db        788 --------------------GlyProThrGlyAlaArgGlyAlaProGlyAspArgGly 800

Qy       1053 TCTCCTTCAGCACCTGCA--------------------CCTGGCTCTGGGCCGCAGCGA 1015
Db        801 GluProGlyProProGlyProAlaGlyPheAlaGlyProProGlyAlaAspGlyGlnPro 820

Qy       1014 CCCTCCTGCGGAACTCCTGGAGCCGAGACCGCTCGCCAGCATCCTGGAGCTCCTTGCCCT 955
Db        821 GlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGlyAspAlaGlyProProGlyPro 840

Qy        954 CGAGCTCCCGCA--------------------------GCTGCCTCTGGCCTTCACTCA 922
Db        841 AlaGlyProAlaGlyProProGlyProIleGlyAsnValGlyAlaProGlyAlaLysGly 860

Qy        921 GCTCGGCCCGCACCTGCTCTGCCTCCTGCTCCAGCTCCCGGATACGCTGGCTGTGCTGGC 862
Db        861 AlaArg---GlySerAlaGlyProProGly---AlaThrGlyPheProGlyAlaAlaGly 878

Qy        861 GGTTCAGGGCCTGAGCTGCCTTTCCTGTGCGGACCAGCTCGCCAATAAGCTCCTCCTTCA 802
Db        879 ArgValGlyProPro--------------GlyProSerGlyAsnAlaGlyProProGly 893

Qy        801 TGCGGATGTTGATAGCCAGCTCCC---GGATCTTCTGCTGGGCCTGGGCCAGCCGCCACT 745
Db        894 Pro------------ProGlyProAlaGlyLysGluGlyGlyLysGlyProArgGly--- 908

Qy        744 CTGAGGCTGTGGCAGGGGGGACCTGGCGGG------CCTGAACTCGGGCCTTGCTCCCAC 691
Db        909 -----------GluThrGlyProAlaGlyArgProGlyGluValGlyPro--------- 921

Qy        690 CAACTGCTCTGGACCCTGGAATGGCTGCATCCAACTCCTCAAGGCAAAGCT--------- 640
Db        922 ------------------------------ProGlyProProGlyProAlaGlyGluLys 931

Qy        639 ------CTGGGCCCTTCCTCTCTGGCAGACTCCCTGGGCGTGCCCCCGCCCTCTGACTGC 586
Db        932 GlySerProGlyAlaAspGlyProAlaGlyAlaProGlyThrProGlyProGlnGly--- 950

Qy        585 AGTTGCTGATCCTATTTCTGCGCAGGTGTAAGGTCCGCCTGGGCGGCTCCTCCT------ 532
Db        951 -----------------IleAlaGlyGlnArgGlyValValGlyLeuProGlyGlnArg 964

Qy        531 --------------CCTCCTCTTCCT--------------CCTCTGAAGCAGCTGAAG 502
Db        965 GlyGluArgGlyPheProGlyLeuProGlyProSerGlyGluProGlyLysGln------ 982

Qy        501 AGCCACTTCCCAGCCTGTTCACCTCAGTCAGCAACT---CAGCTCCAGCCTCCCTGCCAT 445
Db        983 -----------------GlyProSerGlyAlaSerGlyGluArgGlyProProGlyPro 996
```

```
Qy        444 TTGTCACCT---------GCTCTCCCCTCTGCTCAGAGCCAACTTCATCTCCAGGGAGGC 394
Db        997 MetGlyProProGlyLeuAlaGlyProProGlyGluSer-----------------Gly 1010

Qy        393 AGGCAGGCGGCACCA---------------------------------------------- 379
Db       1011 ArgGluGlyAlaProGlyAlaGluGlySerProGlyArgAspGlySerProGlyAlaLys 1030

Qy        378 -----------------TGCCCAGCACATGGGCGTGGGCACCCC------CCAGGGGGG 343
Db       1031 GlyAspArgGlyGluThrGlyProAlaGlyProProGlyAlaProGlyAlaProGlyAla 1050

Qy        342 CTGTATGAGGTC---GAGGCACAAAGGACCCGGGAGGCAGGCCATTCAGGAGCCGCAGGC 286
Db       1051 ProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluThrGlyProAlaGly 1070

Qy        285 CCCCCCAGCCTGGCCGCACCA---------------------------------------- 265
Db       1071 ProAlaGlyProValGlyProValGlyAlaArgGlyProAlaGlyProGlnGlyProArg 1090

Qy        264 ------------------------------------------------------GCTCTA 259
Db       1091 GlyAspLysGlyGluThrGlyGluGlnGlyAspArgGlyIleLysGlyHisArgGlyPhe 1110

Qy        258 ACCGCAGCCGCAGTTCCACCATCTCCTCCTGCTGCTCACGCAGCCGGTCGCTCTGCAGTT 199
Db       1111 SerGlyLeuGlnGlyProProGlyProProGlySerProGlyGluGlnGlyProSerGly 1130

Qy        198 TGTACTGCTCCATGGCGTCCTCCAGCGCAGCCAGAAAGTCTCGGTTCT------------ 151
Db       1131 AlaSerGlyProAlaGlyProArgGlyProProGlySerAlaGlyAlaProGlyLysAsp 1150

Qy        150 ----------------CCTCCTCCAGCCGCGCCACCTGGTTCTGCAGGG---TCAGCAGCT 109
Db       1151 GlyLeuAsnGlyLeuProGlyProIleGlyProProGlyProArgGlyArgThrGlyAsp 1170

Qy        108 GCTGCGCCCCCCTCATCCTCCTTTCGCCCGCCGGCCCCCTGCGCCGCCTGGTCCT 55
Db       1171 AlaGlyProValGlyProProGlyProProGlyProProGlyProProGlyProGlyPro 1188
```

RESULT 12
Q8N473
ID   Q8N473   PRELIMINARY;   PRT; 1464 AA.
AC   Q8N473;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC036531; AAH36531.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Collagen; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Hypothetical protein; Collagen.
SQ   SEQUENCE   1464 AA; 139011 MW; B0581F8D1C89DDE8 CRC64;

Alignment Scores:
Pred. No.: 1.84e-17 Length: 1464

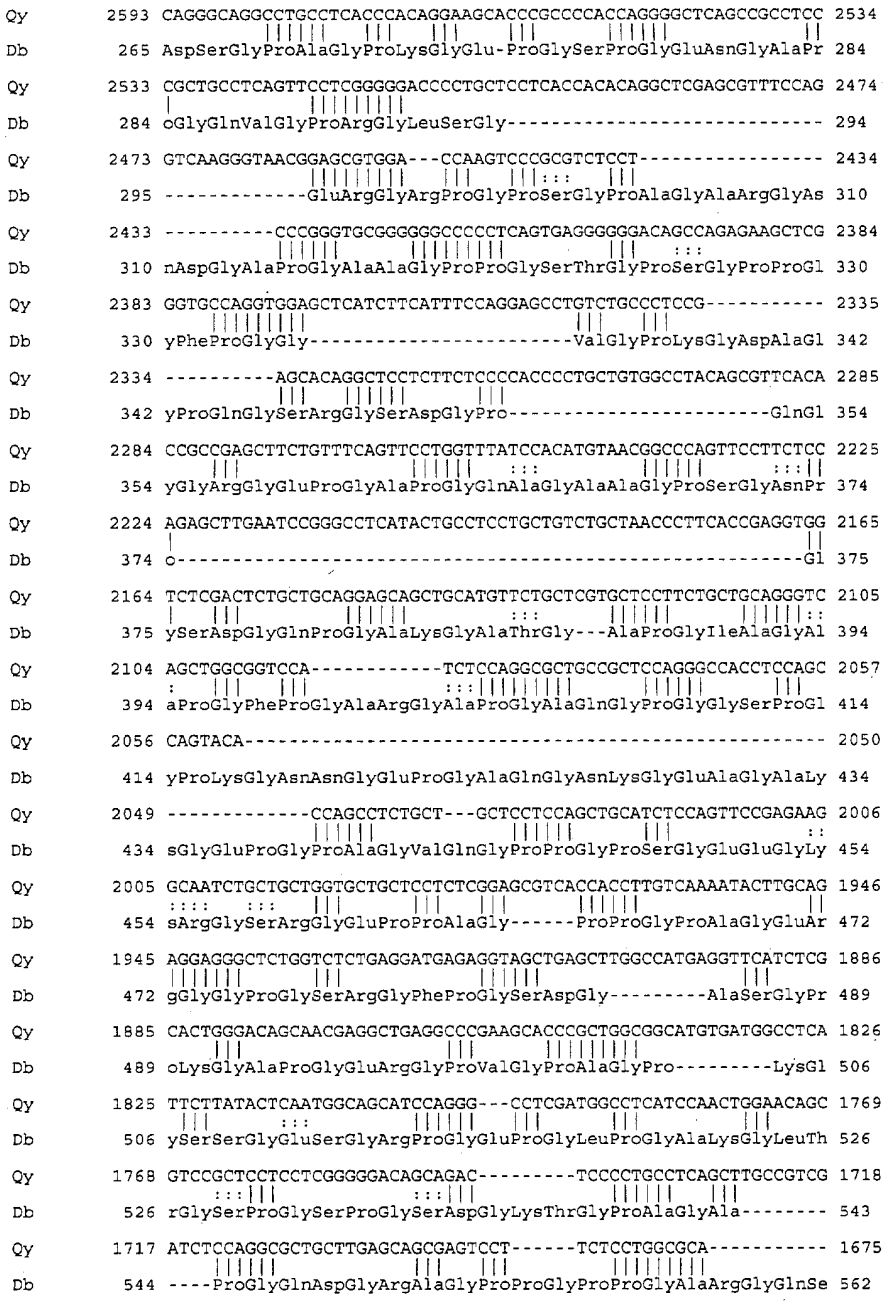
Score:	540.50	Matches:	290
Percent Similarity:	30.49%	Conservative:	57
Best Local Similarity:	25.48%	Mismatches:	424
Query Match:	9.46%	Indels:	367
DB:	4	Gaps:	58

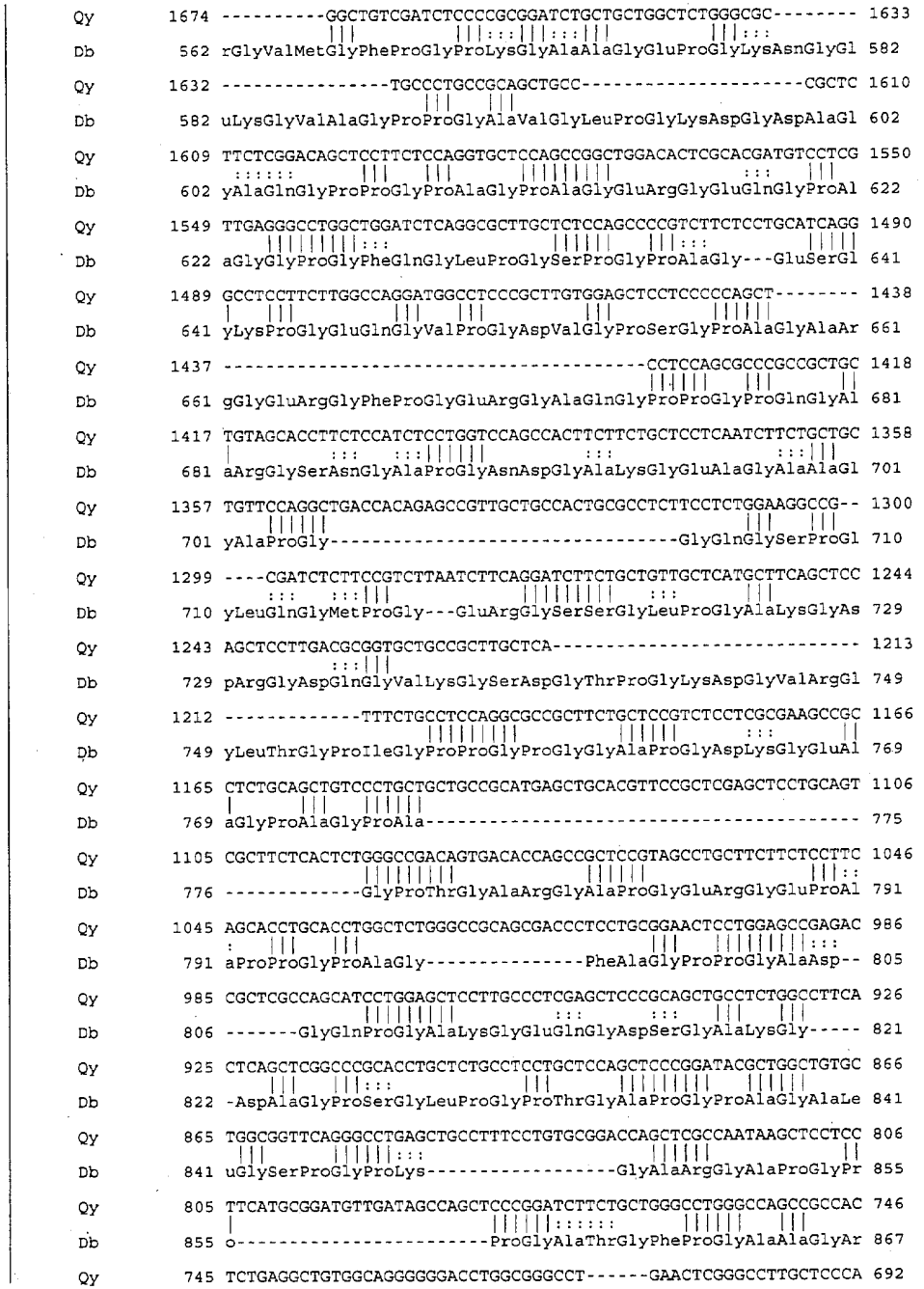
US-10-032-996-6 (1-3121) x Q8N473 (1-1464)

```
Qy       2847 TCCAGTTCTTTTGGGCCATGGCCCAAATTTGTTGATCCCAGTGAGGGTACAGATGAGGGC 2788
Db        214 SerGlyProMetGlyProArgGlyProProGlyProProGlyLysAsnGlyAspAspGly 233

Qy       2787 CTGGATTTAGGGTGTGCGGTAAGGACTGCCCTTCACAGAAGCAAAACAGGCAGCTGCCCC 2728
Db        234 ---GluAlaGlyLysProGlyArgProGly---GluArgGlyProProGlyPro------ 249

Qy       2727 TTTCAGCAGGCTCGGAGTCTCCCTCCAAGGCAGGGTCTGCCCCGAGGGCTTACAGGG--- 2671
Db        250 ---GlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeuPro---GlyMetLysGlyHis 267

Qy       2670 ---GGTTTTTCCGGACATCAATCATCCCCGGGCTGGCTCGTCGCAGTTCCCGCCGGGGGCT 2614
Db        268 ArgGlyPheSerGlyLeuAspGlyAlaLysGly------AspAlaGlyProAlaGly--- 284

Qy       2613 TGGACAAAGGCCCAAAGTTCCAGGGCAGGCCTGCCTCACCCACAGGAAGCACCCGCCCCA 2554
Db        285 ----------------------------------------ProLysGlyGluProGlySer 291

Qy       2553 CCAGGGGCTCAGCCGCCTCCCGCTGCCTCAGTTCCTCGGGGGACCCCT------------ 2506
Db        292 ProGlyGluAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyGluArgGly 311

Qy       2505 ---------GCTCCTCACCACACAGGCTCGAGCGTTTCCAGGTCAAGGGTAACGGAGCGT 2455
Db        312 ArgProGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaAla 331

Qy       2454 GGACCA------------AGTCCCGCGTCTCCTCCCGGGTGCGGGGGGCCCCCTCAGTGA 2407
Db        332 GlyProProGlyProThrGlyProAlaGlyProProGlyPheProGlyAlaValGlyAla 351

Qy       2406 GGGGGGACAGCC---------------AGAGAAGCTCGGGTGCCAGGTGGAGCTCATCTT 2362
Db        352 LysGlyGluAlaGlyProGlnGlyProArgGlySerGluGlyProGlnGlyValArg--- 370

Qy       2361 CATTTCCAGGAGCCTGTCTGCCCTCCGAGCACAGGCTCCTCTTCTCCC------CACCCC 2308
Db        371 ------GlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnPro 388

Qy       2307 TGCTGTGGCCTACAGCGTTCACACCGCCGAGCTTCT---------------GTTTCAGTT 2263
Db        389 GlyAlaAspGlyGlnProGlyAlaLysGlyAlaAsnGlyAlaProGlyIleAlaGlyAla 408

Qy       2262 CCTGGTTTATCCACATGTAACGGCCCA-----------------------GTTCCTTCT 2227
Db        409 ProGlyPheProGlyAlaArgGlyProSerGlyProGlnGlyProGlyGlyProProGly 428

Qy       2226 CCAGAGCTTGAATCCGGGCCTCATACTGCCTCCTGCTGTCTGCTAACCCTTCACCGAGGT 2167
Db        429 ProLysGlyAsnSerGlyGluProGlyAlaPro--------------------------- 439

Qy       2166 GGTCTCGACTCTGCTGCAGGAGCAGCTGCATGTTCTGCTCGTGCTCCTTCTGCTGCAGGG 2107
Db        440 GlySerLysGlyAspThrGlyAlaLysGly------------GluProGlyProValGly 455

Qy       2106 TCAGCTGGCGGTCCATCTCCAGGC---------------------------------- 2083
Db        456 ValGlnGlyProProGlyProAlaGlyGluGluGlyLysArgGlyAlaArgGlyGluPro 475

Qy       2082 ------GCTGCCGCTCCAGGGCCACCT---CCAGCCAGTACACCAGCCTCTGCTGCTCCT 2032
Db        476 GlyProThrGlyLeuProGlyProProGlyGluArgGlyGlyProGlySerArgGlyPhe 495

Qy       2031 CCAGCTGCA--------------------TCTCCAGTTCCGAGAAGGCAATCTGCTGCT 1993
```

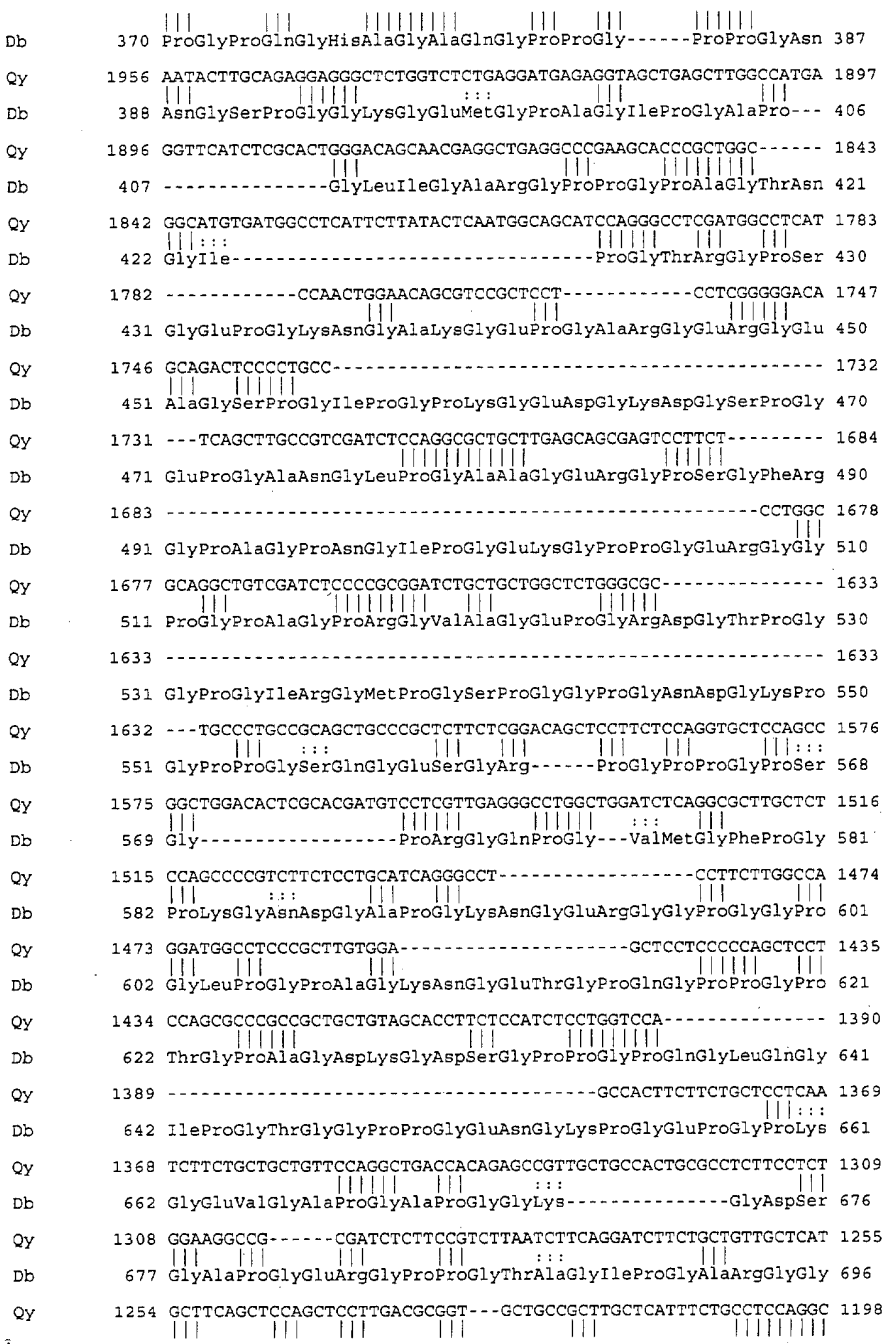
```
Db        496 ProGlyAlaAspGlyValAlaGlyProLysGlyProAlaGlyGluArgGlySerProGly 515

Qy       1992 GGTGCTGCTCCTCTCGGAGCGTCACCACCTTGTCAAAATACTTGCAGAGGAGGGCTCTGG 1933
Db        516 ProAlaGlyProLysGlySerProGlyGluAlaGlyArgProGlyGluAlaGlyLeuPro 535

Qy       1932 TCTCTGAGG---ATGAGAGGTAGCTGAGCTTGGCCATGAGGTTCATCTCGCACTGGGACA 1876
Db        536 GlyAlaLysGlyLeuThrGlySerProGlySerProGlyProAspGlyLysThrGly--- 554

Qy       1875 GCAACGAGGCTGAGGCCCGAAGCACCCGCTGGCGGCATGTGATGGCCTCATTCTTATACT 1816
Db        555 ---------------ProProGlyProAlaGly--------------------------- 560

Qy       1815 CAATGGCAGCATCCAGGGCCTCGATGGCCTCATCCAACTGGAACAGCGTCCGCTCCT--- 1759
Db        561 GlnAspGlyArgProGlyPro------ProGlyProProGlyAlaArgGlyGlnAlaGly 578

Qy       1758 -----------------CCTCGGGGGACAGCAGACTCCCCTGCCTCAGCTTGCCGTCGA 1717
Db        579 ValMetGlyPheProGlyProLysGlyAlaAlaGlyGluProGlyLysAlaGlyGluArg 598

Qy       1716 TCT-----------CCAGGCGCTGCTTGAGCAGCGAGTCCTTCTCCTGGCGCAGGCTGT 1669
Db        599 GlyValProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGluAlaGlyAla 618

Qy       1668 CGATCTCCCCGCGGATCTGCTGCTGGCTCTGGGCGCTGCCCTGCCGCAGCTGCCCGCTCT 1609
Db        619 GlnGlyPro---------------------------------------------------- 621

Qy       1608 TCTCGGACAGCTCCTTCTCCAGGTGCTCCAGCCGGCTGGACACTCGCACGATGTCCTCGT 1549
Db        622 -----------ProGlyProAlaGlyProAlaGlyGluArgGlyGluGlnGlyProAla 637

Qy       1548 TGAGGGCCTGGCTGGATCTCAGGCGCTTGCTCTCCAGCCCCGTCTTCTCCTGCATCAGGG 1489
Db        638 GlySerProGlyPhe---GlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGly 656

Qy       1488 CCTCCTTCTTGGCCAGGATGGCCTCCCGCTTGTGGAGCTCCTCCCCCAGCT--------- 1438
Db        657 LysProGlyGluGlnGlyValProGlyAspLeuGlyAlaProGlyProSerGlyAlaArg 676

Qy       1437 ----------------------------------------CCTCCAGCGCCCGCCGCTGCT 1417
Db        677 GlyGluArgGlyPheProGlyGluArgGlyValGlnGlyProProGlyProAlaGlyPro 696

Qy       1416 GTAGCACCTTCTCCATCTCCTGGTCCAGCCACTTCTTCTGCTCCTCAATCTTCTGCTGCT 1357
Db        697 ArgGlyAlaAsnGlyAlaProGly---------AsnAspGlyAlaLysGlyAspAlaGly 713

Qy       1356 GTTCCAGGCTGACCACAGAGCCGTTGCTGCCACTGCGCCTCTTCCTCTGGAAGGCCG--- 1300
Db        714 AlaProGlyAlaPro-----------------------GlySerGlnGlyAlaProGly 725

Qy       1299 ---CGATCTCTTCCGTCTTAATCTTCAGGATCTTCTGCTGTTGCTCATGCTTCAGCTCCA 1243
Db        726 LeuGlnGlyMetPro------------GlyGluArgGlyAlaAlaGlyLeuProGlyPro 741

Qy       1242 GCTCCTTGACGCGGTGCTGCCGCTTGCTCA---------------------------- 1213
Db        742 LysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGly 761

Qy       1212 --------------------TTTCTGCCTCCAGGCGCCGCTTCTGCTCCGTCTCCTCGC 1174
Db        762 ValArgGlyLeuThrGlyProIleGlyProProGlyProAlaGlyAlaProGlyAspLys 781

Qy       1173 GAAGCCGCCTCTGCAGCTGTCCCTGCTGCTGCCGCATGAGCTGCACGTTCCGCTCGAGCT 1114
Db        782 GlyGluSerGlyProSerGlyProAla------------------------------- 790

Qy       1113 CCTGCAGTCGCTTCTCACTCTGGGCCGACAGTGACACCAGCCGCTCCGTAGCCTGCTTCT 1054
```

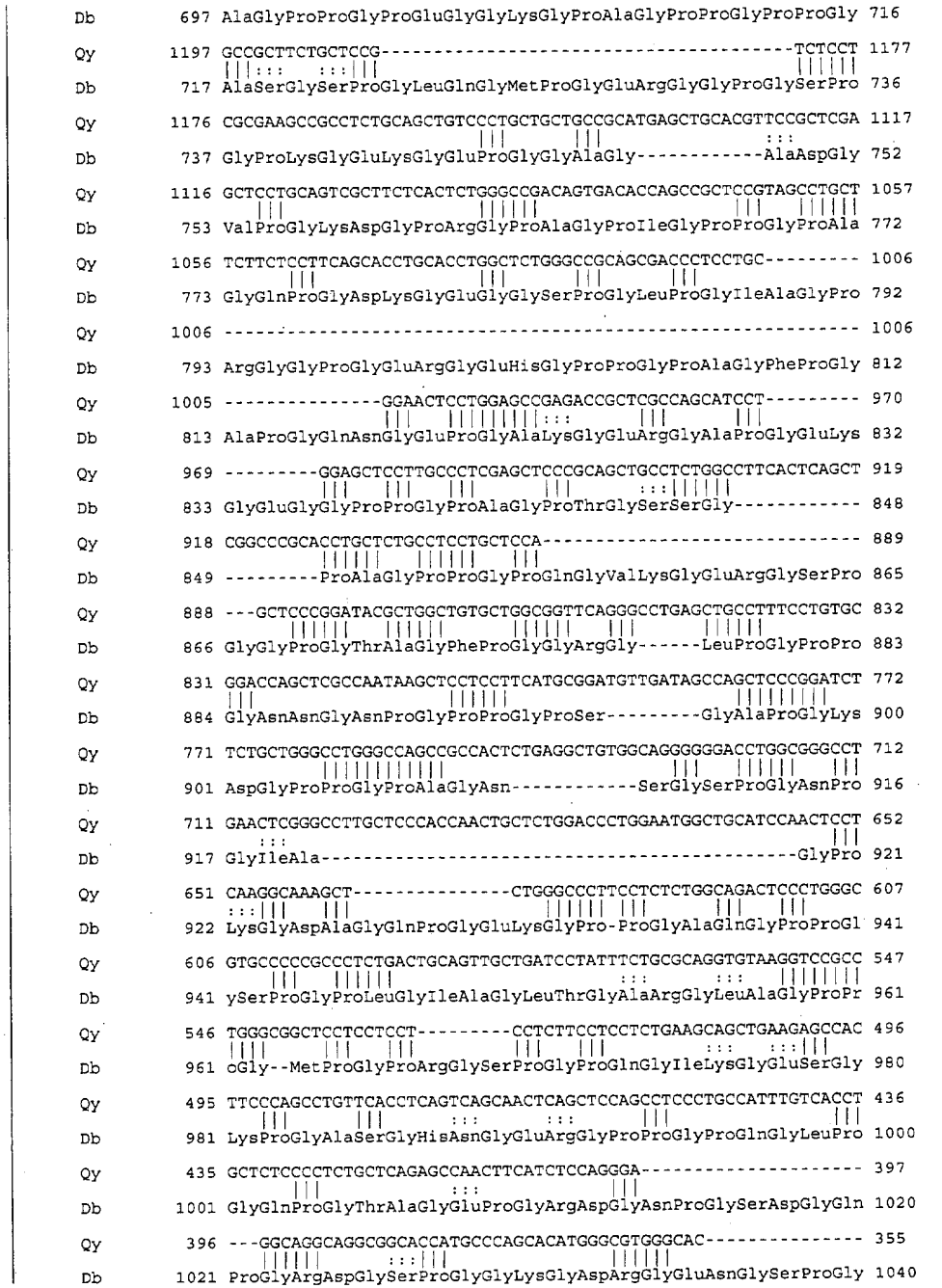
```
Db        791 ----------------------GlyProThrGlyAlaArgGlyAlaProGlyAspArgGly 803

Qy       1053 TCTCCTTCAGCACCTGCA--------------------CCTGGCTCTGGGCCGCAGCGA 1015
              |||       |||                       |||||:::      |||
Db        804 GluProGlyProProGlyProAlaGlyPheAlaGlyProProGlyAlaAspGlyGlnPro 823

Qy       1014 CCCTCCTGCGGAACTCCTGGAGCCGAGACCGCTCGCCAGCATCCTGGAGCTCCTTGCCCT 955
                 :::   |||   ||||||          |||:::          |||   |||   |||
Db        824 GlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGlyAspAlaGlyProProGlyPro 843

Qy        954 CGAGCTCCCGCA--------------------------GCTGCCTCTGGCCTTCACTCA 922
                 ||||||                          |||   |||
Db        844 AlaGlyProAlaGlyProProGlyProIleGlyAsnValGlyAlaProGlyAlaLysGly 863

Qy        921 GCTCGGCCCGCACCTGCTCTGCCTCCTGCTCCAGCTCCCGGATACGCTGGCTGTGCTGGC 862
              ||||||        |||   ||||||       |||   |||:::   |||   ||||||
Db        864 AlaArg---GlySerAlaGlyProProGly---AlaThrGlyPheProGlyAlaAlaGly 881

Qy        861 GGTTCAGGGCCTGAGCTGCCTTTCCTGTGCGGACCAGCTCGCCAATAAGCTCCTCCTTCA 802
                 ||||||                    ||||||:::            ||||||
Db        882 ArgValGlyProPro--------------GlyProSerGlyAsnAlaGlyProProGly 896

Qy        801 TGCGGATGTTGATAGCCAGCTCCC---GGATCTTCTGCTGGGCCTGGGCCAGCCGCCACT 745
                    |||   |||   |||           |||   ||||||
Db        897 Pro------------ProGlyProAlaGlyLysGluGlyGlyLysGlyProArgGly--- 911

Qy        744 CTGAGGCTGTGGCAGGGGGGACCTGGCGGG------CCTGAACTCGGGCCTTGCTCCCAC 691
                    :::   ||||||   |||         |||:::||||||
Db        912 -----------GluThrGlyProAlaGlyArgProGlyGluValGlyPro--------- 924

Qy        690 CAACTGCTCTGGACCCTGGAATGGCTGCATCCAACTCCTCAAGGCAAAGCT--------- 640
                                   |||   |||   |||   |||
Db        925 ------------------------------ProGlyProProGlyProAlaGlyGluLys 934

Qy        639 ------CTGGGCCCTTCCTCTCTGGCAGACTCCCTGGGCGTGCCCCCGCCCTCTGACTGC 586
                  |||             |||   :::   |||   |||   |||
Db        935 GlySerProGlyAlaAspGlyProAlaGlyAlaProGlyThrProGlyProGlnGly--- 953

Qy        585 AGTTGCTGATCCTATTTCTGCGCAGGTGTAAGGTCCGCCTGGGCGGCTCCTCCT------ 532
                           ||||||   |||                  |||
Db        954 ------------------IleAlaGlyGlnArgGlyValValGlyLeuProGlyGlnArg 967

Qy        531 ---------------CCTCCTCTTCCT---------------CCTCTGAAGCAGCTGAAG 502
                           |||   |||||                |||   |||||
Db        968 GlyGluArgGlyPheProGlyLeuProGlyProSerGlyGluProGlyLysGln------ 985

Qy        501 AGCCACTTCCCAGCCTGTTCACCTCAGTCAGCAACT---CAGCTCCAGCCTCCCTGCCAT 445
                           |||     |||:::   :::      ||||||
Db        986 ------------------GlyProSerGlyAlaSerGlyGluArgGlyProProGlyPro 999

Qy        444 TTGTCACCT---------GCTCTCCCCTCTGCTCAGAGCCAACTTCATCTCCAGGGAGGC 394
                 :::   |||       |||   |||       :::|||                 |||
Db       1000 MetGlyProProGlyLeuAlaGlyProProGlyGluSer-----------------Gly 1013

Qy        393 AGGCAGGCGGCACCA---------------------------------------------- 379
                 |||:::   ||||||
Db       1014 ArgGluGlyAlaProGlyAlaGluGlySerProGlyArgAspGlySerProGlyAlaLys 1033

Qy        378 ------------------TGCCCAGCACATGGGCGTGGGCACCCC------CCAGGGGGG 343
                               ||||||         |||   |||      ||||||
Db       1034 GlyAspArgGlyGluThrGlyProAlaGlyProProGlyAlaProGlyAlaProGlyAla 1053

Qy        342 CTGTATGAGGTC---GAGGCACAAAGGACCCGGGAGGCAGGCCATTCAGGAGCCGCAGGC 286
                    |||       |||   ::::::   :::   |||   :::|||   ||||||
Db       1054 ProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluThrGlyProAlaGly 1073

Qy        285 CCCCCCAGCCTGGCCGCACCA---------------------------------------- 265
                 |||          |||
Db       1074 ProAlaGlyProValGlyProValGlyAlaArgGlyProAlaGlyProGlnGlyProArg 1093

Qy        264 ----------------------------------------------------GCTCTA 259

Db       1094 GlyAspLysGlyGluThrGlyGluGlnGlyAspArgGlyIleLysGlyHisArgGlyPhe 1113

Qy        258 ACCGCAGCCGCAGTTCCACCATCTCCTCCTGCTGCTCACGCAGCCGGTCGCTCTGCAGTT 199
                    :::             ||||||   ||||||   :::                :::
Db       1114 SerGlyLeuGlnGlyProProGlyProProGlySerProGlyGluGlnGlyProSerGly 1133

Qy        198 TGTACTGCTCCATGGCGTCCTCCAGCGCAGCCAGAAAGTCTCGGTTCT------------ 151
                    :::   |||       |||           |||   |||   |||:::
Db       1134 AlaSerGlyProAlaGlyProArgGlyProProGlySerAlaGlyAlaProGlyLysAsp 1153

Qy        150 --------------CCTCCTCCAGCCGCGCCACCTGGTTCTGCAGGG---TCAGCAGCT 109
                           |||   |||       |||||||||       |||   :::
Db       1154 GlyLeuAsnGlyLeuProGlyProIleGlyProProGlyProArgGlyArgThrGlyAsp 1173

Qy        108 GCTGCGCCCCCCTCATCCTCCTTTCGCCCGCCGGCCCCCTGCGCCGCCTGGTCCT 55
                 |||   |||       ||||||           ||||||   ||||||||||||
Db       1174 AlaGlyProValGlyProProGlyProProGlyProProGlyProProGlyPro 1191

RESULT 13
Q9IB91
ID   Q9IB91      PRELIMINARY;      PRT;  1447 AA.
AC   Q9IB91;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Type I collagen alpha 1.
GN   COL1A1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RT   "Expression and characterization of Xenopus type I collagen alpha 1
RT   (COL1A1) during embryonic development.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034701; BAA94972.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Collagen; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Collagen.
SQ   SEQUENCE   1447 AA;  137446 MW;  A4A6DD2B4158E38B CRC64;

Alignment Scores:
Pred. No.:              4.08e-17        Length:         1447
Score:                  533.00          Matches:        274
Percent Similarity:     31.75%          Conservative:   60
Best Local Similarity:  26.05%          Mismatches:     416
Query Match:            9.33%           Indels:         302
DB:                     13              Gaps:           48

US-10-032-996-6 (1-3121) x Q9IB91 (1-1447)

Qy       2770 GGTAAGGACTGCCCTTCACAGAAGCAAAACAGGCAGCTGCCCCTTTCAGCAGGCTCGGAG 2711
                 |||:::|||       ::   |||       |||
Db        216 GlyGluAspGlyGluAlaGlyLysProGlyArg-------------------------- 226

Qy       2710 TCTCCCTCCAAGGCAGGG---TCTGCCCCGAGGGCTTACAGGGGGTTTTTCCGGACATCA 2654
                    |||   :::   |||         |||:::      ||||||         |||:::
Db        227 ---ProGlyGluArgGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAla 245

Qy       2653 ATCATCCCCGGGGCTGGCTCGTCGCAGTTCCCGCCGGGGCTTGGACAAAGGCCCAAAGTTC 2594
                    :::||||||:::                        |||||||||   |||   |||
Db        246 GlyLeuProGlyMetLysGlyHisArgGlyPheAsnGlyLeuAsp---GlyAlaLysGly 264
```



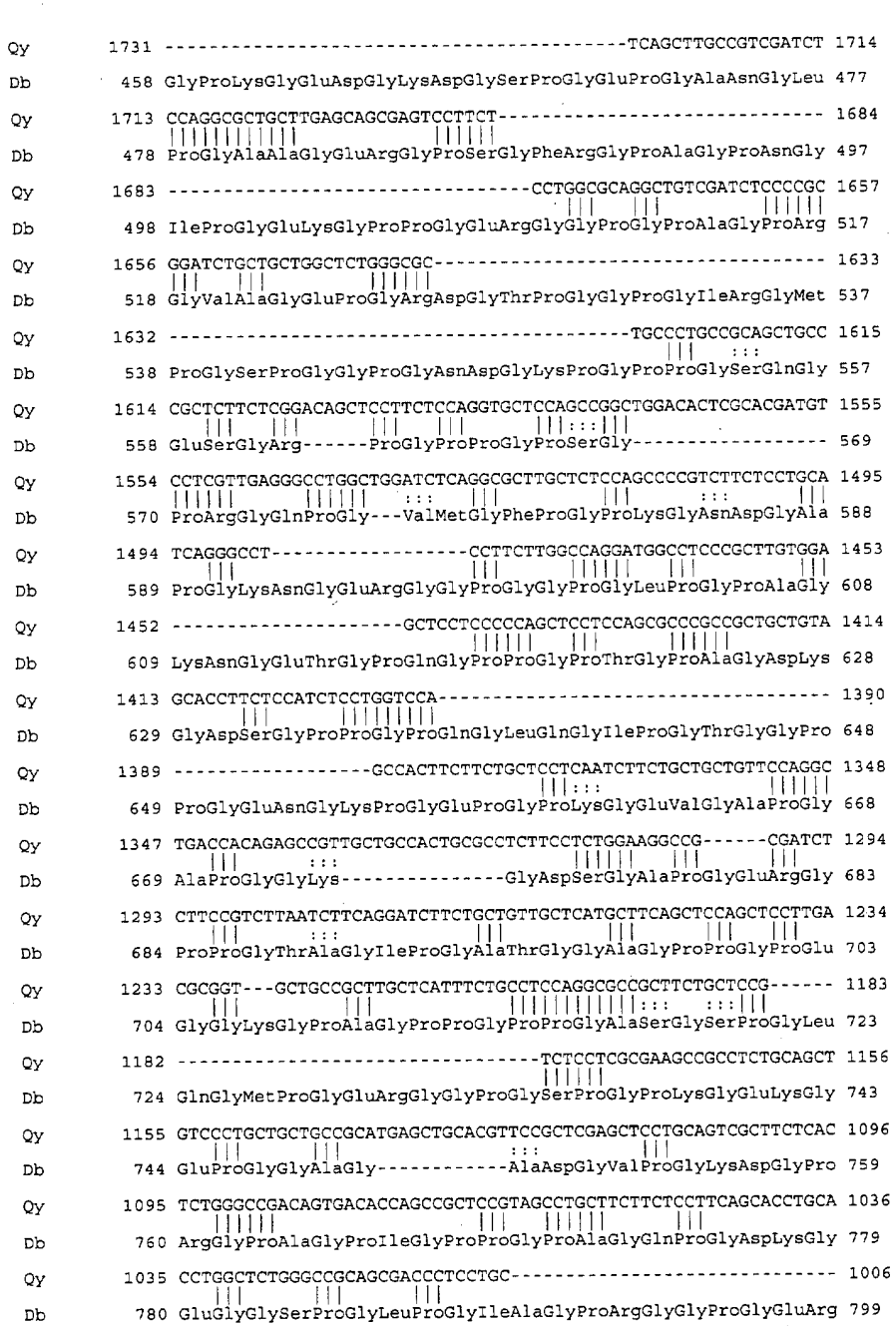


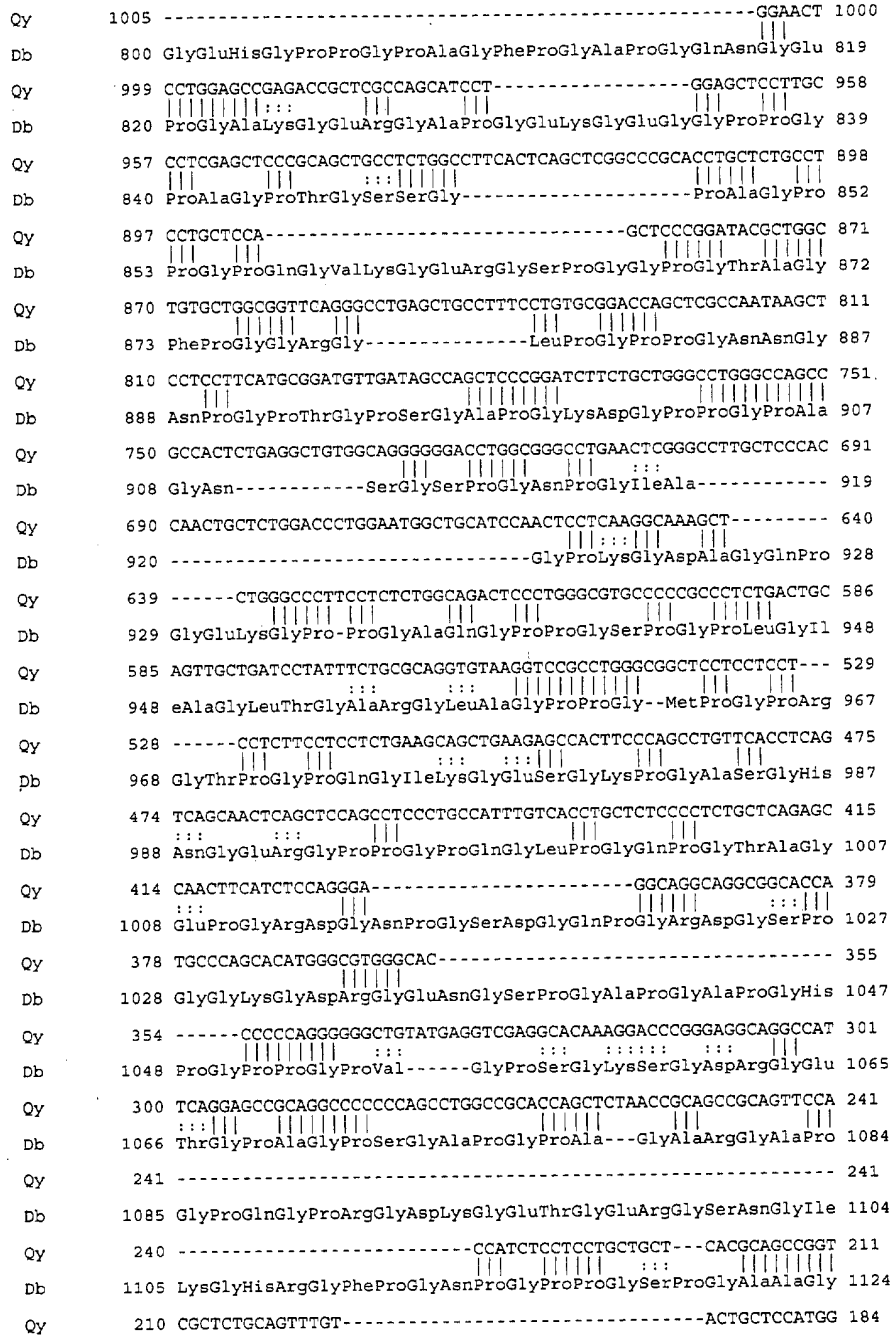
```
Db        867 gVal--------------GlyProProGlyProSerGlyAsnAlaGlyPro-------- 879

Qy        691 CCAACTGCTCTGGACCCTGGAATGGCTGCATCCAACTCCTCAAGGCAAAGCTCTGGGCCC 632

Db        880 ------------------------------ProGlyProSerGlyProAla---GlyLy 888

Qy        631 TTCCTCTCTGGCAGACTCCCTGGGCGTGCCCCCGCCCTCTGACTGCAGTTGCTGATCCTA 572

Db        888 sGluGlyAlaLysGlyProArgGlyGluThrGlyProAlaGlyArgSer-GlyGluProG 908

Qy        571 TTTCTGCGCAGGTGTAAGGTCCGCCTGGGCGGCTCCTCCTCCTCCTCTTCCTCCTCTGAA 512

Db        908 lyAlaAlaGlyProProGlyProProGlyGlu-LysGlySerProGlySerAspGlyPro 927

Qy        511 GCAGCTGAAGAGCCACTTCCCAGCCTGTTCACCTCAGTCAGCAACTCAGCTCCAGCCTCC 452

Db        928 AlaGlyAlaProGlyIleProGlyProGlnGlyValAlaGlySerArgGlyThrValGly 947

Qy        451 CTGCCA--------------------TTTGTCACCTGCTCTCCCCTCTGCTCAGAGCCA 413

Db        948 LeuProGlyMetArgGlyGluArgGlyPheSerGlyLeuProGlyProAlaGlyGluPro 967

Qy        412 ------------ACTTCATCTCCAGGGAGGCAGGCAGGCGGCACCATGCCCAGCA----- 370

Db        968 GlyLysGlnGlySerSerGlyPro-SerGlyGluArgGlyProProGlyProSerGlyPr 987

Qy        369 -CATGGGCGTGGGCACCCCCCAGGGGGGCTGTATGAGGTCGAGGCACAAAGGACCCGGGA 311

Db        987 oProGlyLeuGlyGlyProProGly------------------------------Gl 996

Qy        310 GGCAGGCCATTCAGGAGCCGCAGGCCCCCCCAGCCTGGCCGCACCAGCTCTAACCGCAGC 251

Db        996 uSerGlyArgGluGlyAla------ProGlySerGluGlyAlaProGly---ArgAspGl 1013

Qy        250 CGCAGTTCCACCATCTCCTCCTGCTGCTCACGCAGCCGGTCGCTCTGCAGTTTGTACTGC 191

Db       1013 yAlaValGlyProLysGlyAspArgGlyGluAlaAlaGly-------------------- 1026

Qy        190 TCCATGGCGTCCTCCAGCGCAGCCAGAAAGTCTCGGTTCTCCTCCTCCAGCCGCGCCACC 131

Db       1027 ----------ProProGlyAlaProGlyAlaProGlyAlaProGlyProValGlyProAl 1043

Qy        130 TGGTTCTGCAGGG--------------------TCAGCAGCTGCTGCGCCCCCTCATCC 92

Db       1043 aGlyLysSerGlyAspArgGlyGluThrGlyProSerGlyProAlaGlyProAlaGlyTh 1063

Qy         91 TCCTTTCGCCCGCCGGCCCCCTGCGCCGCCTGGTCCTCGACGGAGGCGCTCTCGATGCCG 32

Db       1063 rAlaGlyAlaArgGlyProAlaGlyProGlnGlyProArgGlyAspLysGlyGluAlaGl 1083

Qy         31 CTATCGGGCCCGGAGGCGGAGCTCAGGG 4

Db       1083 yGluGlnGlyGluArgGlyMetLysGly 1092

RESULT 14
Q8BLW4
ID   Q8BLW4      PRELIMINARY;      PRT;  1464 AA.
AC   Q8BLW4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Collagen alpha 1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK041115; BAC30826.1; -.
SQ   SEQUENCE   1464 AA;  138939 MW;  91F3246D90818449 CRC64;

Alignment Scores:
Pred. No.:              4.31e-17        Length:         1464
Score:                  532.50          Matches:        303
Percent Similarity:     29.30%          Conservative:   48
Best Local Similarity:  25.29%          Mismatches:     452
Query Match:            9.32%           Indels:         395
DB:                     11              Gaps:           58

US-10-032-996-6 (1-3121) x Q8BLW4 (1-1464)

Qy       2709 CTCCCTCCAAGGCAGGGTCTGCCCCGAGGGCTTACAGGG------GGTTTTTCCGGACAT 2656

Db        117 IleProGlyArgAsnGlyAspPro---GlyLeuProGlyGlnProGlyLeuProGlyPro 135

Qy       2655 CAATCATCCCCGGGCTGGCTCGTCGCAGTTCCCGCCGGGGCTTGGACA------------ 2608

Db        136 ProGlySerProGlyIleCysGluSerCysProThrGlyGlyGlnAsnTyrSerProGln 155

Qy       2607 --------------------AAGGCCCAAAGTTCCAGGGCAGGCCTGCCTCACCCACAG 2569

Db        156 PheAspSerTyrAspValLysSerGlyValGlyGlyMetGlyGlyTyrProGlyProAla 175

Qy       2568 GAAGCACCCGCCCCACCAGGGGCTCAGCCGCCTCCCGCTGCCTCAGTTCCTCGGGGGACC 2509

Db        176 GlyProProGlyProProGly---------ProProGlySerSerGlyHisProGlySer 192

Qy       2508 CCT---GCTCCTCACCACACAGGCTCGAGCGTTTCCAGGTCAAGGGTAACGGAGCGTGGA 2452

Db        193 ProGlySerProGlyTyrGlnGlyProPro---------GlyGluProGlyGlnAlaGly 209

Qy       2451 CCAAGT---CCCGCGTCTCCTCCCGGGTGCGGGGGGCCCCCTCAGTGAGGGGGGACAGCC 2395

Db        210 ProAlaGlyProProGlyProProGlyAlaLeuGlyProAlaGlyProAlaGlyLysAsp 229

Qy       2394 AGAGAAGCTCGGGTGCCAGGTGGAGCTCATCTTCATTTCCAGGAGCCTGTCTGCCCTCCG 2335

Db        230 GlyGluSerGlyArgProGlyArgPro------------GlyGluArgGlyLeuProGly 245

Qy       2334 AGCACAGGCTCCTCTTCTCCCCACCCCTGCTGTGGCCTA------CAGCGTTCACACCGC 2281

Db        246 ProProGlyIleLysGlyProAlaGlyMetProGlyPheProGlyMetLysGlyHisArg 265

Qy       2280 ------------------------CGAGCTTCTGTTTCAGTTCCTGGTTTATCCACATGT 2245

Db        266 GlyPheAspGlyArgAsnGlyGluLysGlyGluThrGlyAlaProGlyLeuLysGlyGlu 285

Qy       2244 AACGGC--------------CCAGTTCCTTCTCCAGAGCTTGAATCCGGGCCTCATACT 2200

Db        286 AsnGlyLeuProGlyAspAsnGlyAlaPro***Pro---------MetGlyProArgGly 302

Qy       2199 GCCTCCTGCTGTCTGCTAACCCTTCACCGAGGT------GGTCTCGACTCTGCTGCAGGA 2146

Db        303 AlaProGly--------------GluArgGlyArgProGlyLeuProGlyAlaAlaGly 317

Qy       2145 GCAGCTGCATGTTCTGCTCGTGCTCCTTCTGCTGCAGGGTCAGCTGGCGGTCCATCTCCA 2086

Db        318 AlaArgGly------------AsnAspGlyAlaArgGlySerGlyGlyGlnProGlyPro 333

Qy       2085 GGCGCTGCCGCTCCAGGGCCACCT---------------------------------- 2062

Db        334 ------------ProGlyProProGlyThrAlaGlyPheProGlySerProGlyAlaLys 349

Qy       2061 ------------CCAGCCAGTACACCAGCCTCTGCTGCTCCTCCAGCTGCATCTCCAGTT 2014

Db        350 GlyGluValGlyProAlaGlySerProGlySerAsnGlySerProGlyGlnArgGlyGlu 369

Qy       2013 CCGAGAAGGCAATCTGCTGCTGGTGCT---GCTCCTCTCGGAGCGTCACCACCTTGTCAA 1957
```


Db 370 ProGlyProGlnGlyHisAlaGlyAlaGlnGlyProProGly------ProProGlyAsn 387
Qy 1956 AATACTTGCAGAGGAGGGCTCTGGTCTCTGAGGATGAGAGGTAGCTGAGCTTGGCCATGA 1897
Db 388 AsnGlySerProGlyGlyLysGlyGluMetGlyProAlaGlyIleProGlyAlaPro--- 406
Qy 1896 GGTTCATCTCGCACTGGGACAGCAACGAGGCTGAGGCCCGAAGCACCCGCTGGC------ 1843
Db 407 --------------GlyLeuIleGlyAlaArgGlyProProGlyProAlaGlyThrAsn 421
Qy 1842 GGCATGTGATGGCCTCATTCTTATACTCAATGGCAGCATCCAGGGCCTCGATGGCCTCAT 1783
Db 422 GlyIle--------------------------------ProGlyThrArgGlyProSer 430
Qy 1782 ------------CCAACTGGAACAGCGTCCGCTCCT------------CCTCGGGGGACA 1747
Db 431 GlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyAlaArgGlyGluArgGlyGlu 450
Qy 1746 GCAGACTCCCCTGCC-- 1732
Db 451 AlaGlySerProGlyIleProGlyProLysGlyGluAspGlyLysAspGlySerProGly 470
Qy 1731 ---TCAGCTTGCCGTCGATCTCCAGGCGCTGCTTGAGCAGCGAGTCCTTCT--------- 1684
Db 471 GluProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyProSerGlyPheArg 490
Qy 1683 --CCTGGC 1678
Db 491 GlyProAlaGlyProAsnGlyIleProGlyGluLysGlyProProGlyGluArgGlyGly 510
Qy 1677 GCAGGCTGTCGATCTCCCCGCGGATCTGCTGCTGGCTCTGGGCGC--------------- 1633
Db 511 ProGlyProAlaGlyProArgGlyValAlaGlyGluProGlyArgAspGlyThrProGly 530
Qy 1633 -- 1633
Db 531 GlyProGlyIleArgGlyMetProGlySerProGlyGlyProGlyAsnAspGlyLysPro 550
Qy 1632 ---TGCCCTGCCGCAGCTGCCCGCTCTTCTCGGACAGCTCCTTCTCCAGGTGCTCCAGCC 1576
Db 551 GlyProProGlySerGlnGlyGluSerGlyArg------ProGlyProProGlyProSer 568
Qy 1575 GGCTGGACACTCGCACGATGTCCTCGTTGAGGGCCTGGCTGGATCTCAGGCGCTTGCTCT 1516
Db 569 Gly-----------------ProArgGlyGlnProGly---ValMetGlyPheProGly 581
Qy 1515 CCAGCCCCGTCTTCTCCTGCATCAGGGCCT-----------------CCTTCTTGGCCA 1474
Db 582 ProLysGlyAsnAspGlyAlaProGlyLysAsnGlyGluArgGlyGlyProGlyGlyPro 601
Qy 1473 GGATGGCCTCCCGCTTGTGGA--------------------GCTCCTCCCCCAGCTCCT 1435
Db 602 GlyLeuProGlyProAlaGlyLysAsnGlyGluThrGlyProGlnGlyProProGlyPro 621
Qy 1434 CCAGCGCCCGCCGCTGCTGTAGCACCTTCTCCATCTCCTGGTCCA-------------- 1390
Db 622 ThrGlyProAlaGlyAspLysGlyAspSerGlyProProGlyProGlnGlyLeuGlnGly 641
Qy 1389 ------------------------------------GCCACTTCTTCTGCTCCTCAA 1369
Db 642 IleProGlyThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLys 661
Qy 1368 TCTTCTGCTGCTGTTCCAGGCTGACCACAGAGCCGTTGCTGCCACTGCGCCTCTTCCTCT 1309
Db 662 GlyGluValGlyAlaProGlyAlaProGlyGlyLys--------------GlyAspSer 676
Qy 1308 GGAAGGCCG------CGATCTCTTCCGTCTTAATCTTCAGGATCTTCTGCTGTTGCTCAT 1255
Db 677 GlyAlaProGlyGluArgGlyProProGlyThrAlaGlyIleProGlyAlaArgGlyGly 696
Qy 1254 GCTTCAGCTCCAGCTCCTTGACGCGGT---GCTGCCGCTTGCTCATTTCTGCCTCCAGGC 1198


Db 697 AlaGlyProProGlyProGluGlyGlyLysGlyProAlaGlyProProGlyProProGly 716
Qy 1197 GCCGCTTCTGCTCCG--TCTCCT 1177
Db 717 AlaSerGlySerProGlyLeuGlnGlyMetProGlyGluArgGlyGlyProGlySerPro 736
Qy 1176 CGCGAAGCCGCCTCTGCAGCTGTCCCTGCTGCTGCCGCATGAGCTGCACGTTCCGCTCGA 1117
Db 737 GlyProLysGlyGluLysGlyGluProGlyGlyAlaGly------------AlaAspGly 752
Qy 1116 GCTCCTGCAGTCGCTTCTCACTCTGGGCCGACAGTGACACCAGCCGCTCCGTAGCCTGCT 1057
Db 753 ValProGlyLysAspGlyProArgGlyProAlaGlyProIleGlyProProGlyProAla 772
Qy 1056 TCTTCTCCTTCAGCACCTGCACCTGGCTCTGGGCCGCAGCGACCCTCCTGC--------- 1006
Db 773 GlyGlnProGlyAspLysGlyGluGlyGlySerProGlyLeuProGlyIleAlaGlyPro 792
Qy 1006 ---------- 1006
Db 793 ArgGlyGlyProGlyGluArgGlyGluHisGlyProProGlyProAlaGlyPheProGly 812
Qy 1005 --------------GGAACTCCTGGAGCCGAGACCGCTCGCCAGCATCCT--------- 970
Db 813 AlaProGlyGlnAsnGlyGluProGlyAlaLysGlyGluArgGlyAlaProGlyGluLys 832
Qy 969 ---------GGAGCTCCTTGCCCTCGAGCTCCCGCAGCTGCCTCTGGCCTTCACTCAGCT 919
Db 833 GlyGluGlyGlyProProGlyProAlaGlyProThrGlySerSerGly------------ 848
Qy 918 CGGCCCGCACCTGCTCTGCCTCCTGCTCCA------------------------------ 889
Db 849 ---------ProAlaGlyProProGlyProGlnGlyValLysGlyGluArgGlySerPro 865
Qy 888 ---GCTCCCGGATACGCTGGCTGTGCTGGCGGTTCAGGGCCTGAGCTGCCTTTCCTGTGC 832
Db 866 GlyGlyProGlyThrAlaGlyPheProGlyGlyArgGly------LeuProGlyProPro 883
Qy 831 GGACCAGCTCGCCAATAAGCTCCTCCTTCATGCGGATGTTGATAGCCAGCTCCCGGATCT 772
Db 884 GlyAsnAsnGlyAsnProGlyProProGlyProSer---------GlyAlaProGlyLys 900
Qy 771 TCTGCTGGGCCTGGGCCAGCCGCCACTCTGAGGCTGTGGCAGGGGGGACCTGGCGGGCCT 712
Db 901 AspGlyProProGlyProAlaGlyAsn-----------SerGlySerProGlyAsnPro 916
Qy 711 GAACTCGGGCCTTGCTCCCACCAACTGCTCTGGACCCTGGAATGGCTGCATCCAACTCCT 652
Db 917 GlyIleAla--GlyPro 921
Qy 651 CAAGGCAAAGCT---------------CTGGGCCCTTCCTCTCTGGCAGACTCCCTGGGC 607
Db 922 LysGlyAspAlaGlyGlnProGlyGluLysGlyPro-ProGlyAlaGlnGlyProProGl 941
Qy 606 GTGCCCCCGCCCTCTGACTGCAGTTGCTGATCCTATTTCTGCGCAGGTGTAAGGTCCGCC 547
Db 941 ySerProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyProPr 961
Qy 546 TGGGCGGCTCCTCCTCCT---------CCTCTTCCTCCTCTGAAGCAGCTGAAGAGCCAC 496
Db 961 oGly--MetProGlyProArgGlySerProGlyProGlnGlyIleLysGlyGluSerGly 980
Qy 495 TTCCCAGCCTGTTCACCTCAGTCAGCAACTCAGCTCCAGCCTCCCTGCCATTTGTCACCT 436
Db 981 LysProGlyAlaSerGlyHisAsnGlyGluArgGlyProProGlyProGlnGlyLeuPro 1000
Qy 435 GCTCTCCCCTCTGCTCAGAGCCAACTTCATCTCCAGGGA---------------------- 397
Db 1001 GlyGlnProGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyGln 1020
Qy 396 ---GGCAGGCAGGCGGCACCATGCCCAGCACATGGGCGTGGGCAC--------------- 355
Db 1021 ProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGluAsnGlySerProGly 1040

```
Qy        354 ----------------------------CCCCCAGGGGGGCTGTATGAGGTCGAGGCACAA 322
Db       1041 AlaProGlyAlaProGlyHisProGlyProProGlyProVal------GlyProSerGly 1058

Qy        321 AGGACCCGGGAGGCAGGCCATTCAGGAGCCGCAGGCCCCCCCAGCCTGGCCGCACCAGCT 262
Db       1059 LysSerGlyAspArgGlyGluThrGlyProAlaGlyProSerGlyAlaProGlyProAla 1078

Qy        261 CTAACCGCAGCCGCAGTTCCA------------------------------------ 241
Db       1079 ---GlyAlaArgGlyAlaProGlyProGlnGlyProArgGlyAspLysGlyGluThrGly 1097

Qy        240 ------------------------------------------------CCATCTCCTCCT 229
Db       1098 GluArgGlySerAsnGlyIleLysGlyHisArgGlyPheProGlyAsnProGlyProPro 1117

Qy        228 GCTGCT---CACGCAGCCGGTCGCTCTGCAGTTTGT----------------------- 196
Db       1118 GlySerProGlyAlaAlaGlyHisGlnGlyAlaIleGlySerProGlyProAlaGlyPro 1137

Qy        195 ---------ACTGCTCCATGGCGTCCTCCAGCGCAGCCAGAAAGTCTCGGTTCTCCTCCT 145
Db       1138 ArgGlyProValGlyProHisGlyProProGlyLysAspGlyThrSerGlyHisProGly 1157

Qy        144 CCAGCCGCGCCACCTGGTTCTGCAGGGTCA--------------GCAGCTGCTGCGCCC 100
Db       1158 ProIleGlyProProGlyProArgGlyAsnArgGlyGluArgGlySerGluGlySerPro 1177

Qy         99 CCTCATCCT------CCTTTCGCCCGCCGGCCCCCTGCGCCGCCTGGTCCTCGAC---GG 49
Db       1178 GlyHisProGlyGlnProGlyProProGlyProProGlyAlaProGlyPro-CysCysGl 1197

Qy         48 AGGCGCTCTCGATGCCGCTATCGGGCCCGGAGGCGGAGCTCAGGGC 3
Db       1197 yGlyGlyAlaAlaAlaIleAlaGlyValGlyGlyGluLysSerGly 1212

RESULT 15
Q8BKY2
ID   Q8BKY2      PRELIMINARY;      PRT;  1464 AA.
AC   Q8BKY2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Collagen alpha 1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK048546; BAC33370.1; -.
SQ   SEQUENCE   1464 AA;  138947 MW;  1E4ED9539EF42C12 CRC64;

Alignment Scores:
Pred. No.:              4.79e-17        Length:         1464
Score:                  531.50          Matches:        299
Percent Similarity:     29.22%          Conservative:   49
Best Local Similarity:  25.10%          Mismatches:     462
Query Match:            9.30%           Indels:         381
DB:                     11              Gaps:           54

US-10-032-996-6 (1-3121) x Q8BKY2 (1-1464)

Qy       2709 CTCCCTCCAAGGCAGGGTCTGCCCCGAGGGCTTACAGGG------GGTTTTTCCGGACAT 2656
Db        117 IleProGlyArgAsnGlyAspPro---GlyLeuProGlyGlnProGlyLeuProGlyPro 135

Qy       2655 CAATCATCCCCGGGCTGGCTCGTCGCAGTTCCCGCCGGGGCTTGGACA------------ 2608
Db        136 ProGlySerProGlyIleCysGluSerCysProThrGlyGlyGlnAsnTyrSerProGln 155

Qy       2607 --------------------AAGGCCCAAAGTTCCAGGGCAGGCCTGCCTCACCCACAG 2569
Db        156 PheAspSerTyrAspValLysSerGlyValGlyGlyMetGlyGlyTyrProGlyProAla 175

Qy       2568 GAAGCACCCGCCCCACCAGGGGCTCAGCCGCCTCCCGCTGCCTCAGTTCCTCGGGGGACC 2509
Db        176 GlyProProGlyProProGly---------ProProGlySerSerGlyHisProGlySer 192

Qy       2508 CCT---GCTCCTCACCACACAGGCTCGAGCGTTTCCAGGTCAAGGGTAACGGAGCGTGGA 2452
Db        193 ProGlySerProGlyTyrGlnGlyProPro---------GlyGluProGlyGlnAlaGly 209

Qy       2451 CCAAGT---CCCGCGTCTCCTCCCGGGTGCGGGGGGCCCCCTCAGTGAGGGGGGACAGCC 2395
Db        210 ProAlaGlyProProGlyProProGlyAlaLeuGlyProAlaGlyProAlaGlyLysAsp 229

Qy       2394 AGAGAAGCTCGGGTGCCAGGTGGAGCTCATCTTCATTTCCAGGAGCCTGTCTGCCCTCCG 2335
Db        230 GlyGluSerGlyArgProGlyArgPro------------GlyGluArgGlyLeuProGly 245

Qy       2334 AGCACAGGCTCCTCTTCTCCCCACCCCTGCTGTGGCCTA------CAGCGTTCACACCGC 2281
Db        246 ProProGlyIleLysGlyProAlaGlyMetProGlyPheProGlyMetLysGlyHisArg 265

Qy       2280 ------------------------CGAGCTTCTGTTTCAGTTCCTGGTTTATCCACATGT 2245
Db        266 GlyPheAspGlyArgAsnGlyGluLysGlyGluThrGlyAlaProGlyLeuLysGlyGlu 285

Qy       2244 AACGGCCCAGTTCCTTCTCCAGAGCTTGAATCCGGGCCTCATACTGCCTCCTGCTGTCTG 2185
Db        286 AsnGly---LeuProGlyAspAsnGlyAlaProGlyProMetGlyProArgGlyAlaPro 304

Qy       2184 CTAACCCTTCACCGAGGTGGTCTCGACTCTGCTGCAGGAGCAGCTGCATGTTCTGCTCGT 2125
Db        305 GlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArgGly------------ 320

Qy       2124 GCTCCTTCTGCTGCAGGGTCAGCTGGCGGTCCATCTCCAGGCGCTGCCGCTCCAGGGCCA 2065
Db        321 LysAspGlyAlaArgGlySerAspGlyGlnProGlyPro------------ProGlyPro 336

Qy       2064 CCT----------------------------------------------------CCAGCCAGT 2053
Db        337 ProGlyThrAlaGlyPheProGlySerProGlyAlaLysGlyGluValGlyProAlaGly 356

Qy       2052 ACACCAGCCTCTGCTGCTCCTCCAGCTGCATCTCCAGTTCCGAGAAGGCAATCTGCTGCT 1993
Db        357 SerProGlySerAsnGlySerProGlyGlnArgGlyGluProGlyProGlyGlnGlyHisAla 376

Qy       1992 GGTGCT---GCTCCTCTCGGAGCGTCACCACCTTGTCAAAATACTTGCAGAGGAGGGCTC 1936
Db        377 GlyAlaGlnGlyProProGly------ProProGlyAsnAsnGlySerProGlyGlyLys 394

Qy       1935 TGGTCTCTGAGGATGAGAGGTAGCTGAGCTTGGCCATGAGGTTCATCTCGCACTGGGACA 1876
Db        395 GlyGluMetGlyProAlaGlyIleProGlyAlaPro-----------------GlyLeu 408

Qy       1875 GCAACGAGGCTGAGGCCCGAAGCACCCGCTGGC------GGCATGTGATGGCCTCATTCT 1822
Db        409 IleGlyAlaArgGlyProProGlyProAlaGlyThrAsnGlyIle---------------- 423

Qy       1821 TATACTCAATGGCAGCATCCAGGGCCTCGATGGCCTCAT------------CCAACTGGA 1774
Db        424 -----------------ProGlyThrArgGlyProSerGlyGluProGlyLysAsnGly 437

Qy       1773 ACAGCGTCCGCTCCT------------CCTCGGGGGACAGCAGACTCCCCTGCC------ 1732
Db        438 AlaLysGlyGluProGlyAlaArgGlyGluArgGlyGluAlaGlySerProGlyIlePro 457
```





```
                                                                      |||
Db       1125 HisGlnGlyAlaIleGlySerProGlyProAlaGlyProArgGlyProValGlyProHis 1144

Qy        183 CGTCCTCCAGCGCAGCCAGAAAGTCTCGGTTCTCCTCCTCCAGCCGCGCCACCTGGTTCT 124
              ||||||   :::      :::   |||   |||   |||       |||||||||
Db       1145 GlyProProGlyLysAspGlyThrSerGlyHisProGlyProIleGlyProProGlyPro 1164

Qy        123 GCAGGGTCA--------------GCAGCTGCTGCGCCCCCTCATCCT------CCTTTC 85
                ||:::               :::      :::|||   ||||||       |||
Db       1165 ArgGlyAsnArgGlyGluArgGlySerGluGlySerProGlyHisProGlyGlnProGly 1184

Qy         84 GCCCGCCGGCCCCCTGCGCCGCCTGGTCCTCGAC---GGAGGCGCTCTCGATGCCGCTAT 28
                        ||||||      |||||||||       ||||||         |||
Db       1185 ProProGlyProProGlyAlaProGlyPro-CysCysGlyGlyGlyAlaAlaAlaIleAl 1204

Qy         27 CGGGCCCGGAGGCGGAGCTCAGGGC 3
              |||   ||||||          |||
Db       1204 aGlyValGlyGlyGluLysSerGly 1212
```

Search completed: January 29, 2004, 23:53:31
Job time : 235 secs